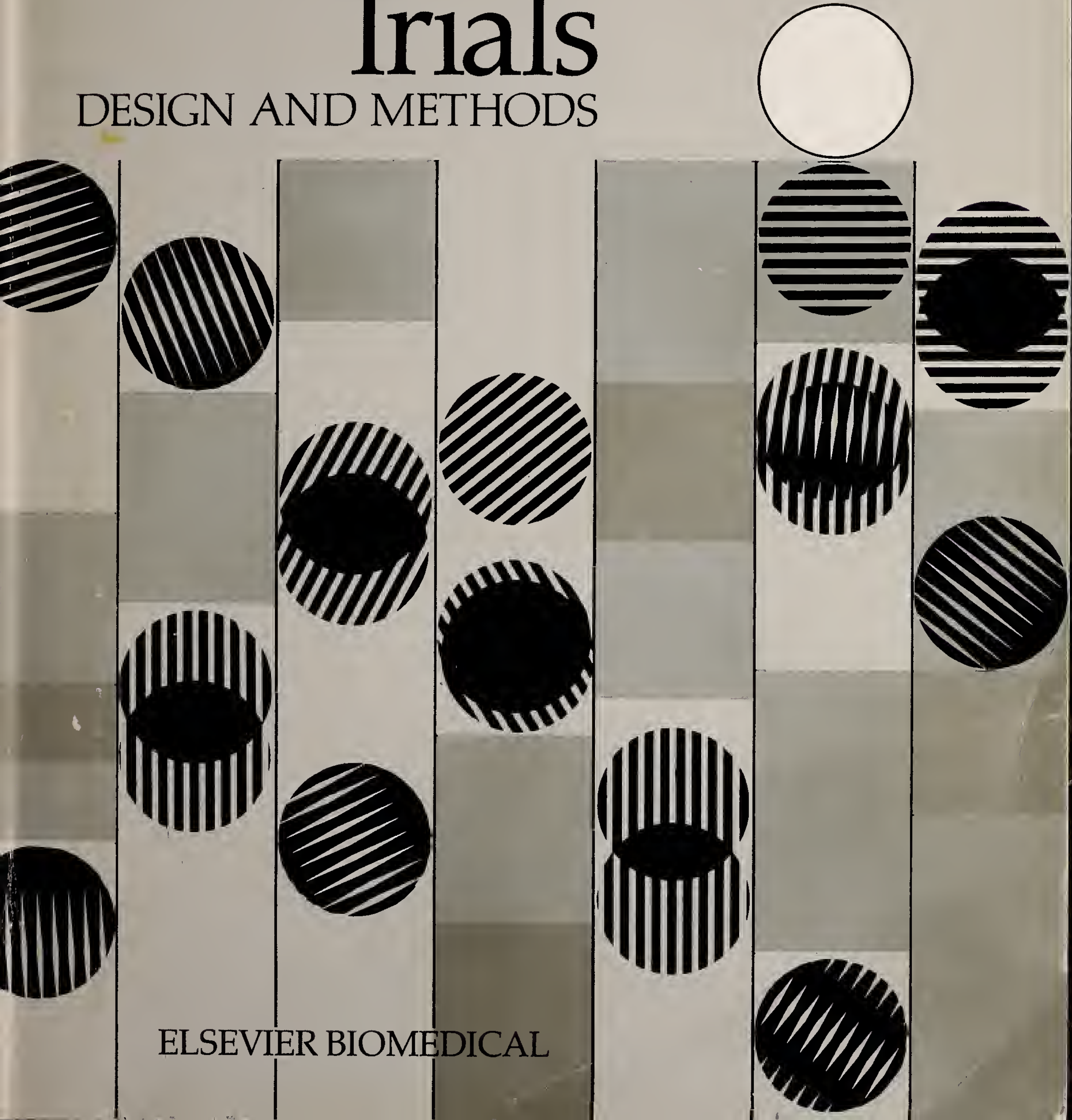
ISSN 0197-2456
CCLTDH 3(3) 163–310 (1982)

Volume 3, Number 3
September 1982

Controlled Clinical Trials

DESIGN AND METHODS



ELSEVIER BIOMEDICAL

Controlled Clinical Trials

DESIGN AND METHODS

Official Journal of the Society for Clinical Trials

Volume 3, Number 3, 1982

ELSEVIER BIOMEDICAL

Controlled Clinical Trials

DESIGN AND METHODS

Official Journal of the Society for Clinical Trials

AIMS AND SCOPE

Controlled Clinical Trials is an international journal designed for the publication of manuscripts which deal with the design, methods and operational aspects of prospective follow-up studies with emphasis on controlled clinical trials. Manuscripts submitted should appeal to a readership drawn from disciplines including medicine, biostatistics, epidemiology, computer science, management science, behavioral science and bioethics. Full-length papers and short communications not exceeding 1,500 words in the following subject areas will be published:

Unusual design features of specific studies which may be applicable to other studies

Organizational structures which facilitate multicenter investigations

General or specific design and operating features for data processing systems used for maintaining data base for long-term follow-up studies

Medical, legal and ethical issues in the conduct of prospective follow-up studies

Management of multicenter investigations

Methods for assessing patient compliance in ongoing trials

Design principles for forms and other data collection techniques for long-term follow-up studies

Quality assurance procedures for error detection in the data generation and processing procedures in long-term follow-up studies

Methods of sample size calculation designed for long-term trials

Historical review of clinical trials

Methods of monitoring clinical trials for evidence of adverse or beneficial treatment effects

Cost assessment of procedures used for data generation and processing in clinical trials

Assessment of the impact of clinical trials on the practice of medicine

Findings from a specific trial or follow-up study which illustrate specific methodological points

Editor-in-Chief

Curtis L. Meinert
Department of Epidemiology
Johns Hopkins School of Hygiene and Public Health
615 North Wolfe Street, Baltimore, Maryland 21205, USA

Associate Editors

Niels Tygstrup
Medical Department A
Rigshospitalet
9 Blegdamsvej
DK-2100 Copenhagen, Denmark

O. Dale Williams
Department of Biostatistics
School of Public Health
University of North Carolina
Chapel Hill, North Carolina 27514, USA

Assistant Editor

Barbara S. Hawkins
Department of Epidemiology
John Hopkins School of Hygiene
and Public Health
615 North Wolfe Street
Baltimore, Maryland 21205

Book Review Editor

R. Sylvester
Institut Jules Bordet
Rue Héger—Bordet 1
1000 Brussels, Belgium

Editorial Board

H. Abe (1984)
Osaka, Japan

J. Bearman (1983)
Beersheva, Israel

K. Breddin (1982)
Frankfurt, Germany

B. Wm. Brown, Jr. (1982)
Stanford, California, USA

P.L. Canner (1983)
Baltimore, Maryland, USA

T.C. Chalmers (1984)
New York, New York, USA

H.O. Conn (1983)
West Haven, Connecticut, USA

R. Flamant (1984)
Villejuif, France

L.C. Gatewood (1982)
Minneapolis, Minnesota, USA

R.S. Gordon (1984)
Bethesda, Maryland, USA

G.B. Hill (1982)
Ottawa, Canada

C.R. Klimt (1983)
Baltimore, Maryland, USA

K. Lebacqz (1984)
Berkeley, California, USA

J. Lubsen (1983)
Rotterdam, The Netherlands

T.E. Prout (1982)
Baltimore, Maryland, USA

M. Rozencweig (1983)
Brussels, Belgium

C.L. Rümke (1982)
Amsterdam, The Netherlands

R.B. Shekelle (1984)
Chicago, Illinois, USA

P.G. Smith (1983)
London, UK

S. Tominaga (1984)
Nagoya, Japan

Controlled Clinical Trials is published quarterly, in one volume, by Elsevier Science Publishing Co., Inc. Subscription price is $80.00 per year. Foreign subscribers: please add $6.00 for postage and handling for surface mail. For air mail within the U.S.A. add $8.00. Air mail to Europe and Canada: $16.00; all other countries: $20.00. Claims for missing issues can be honored only up to three months for domestic addresses, and six months for foreign addresses. Duplicate copies will not be sent to replace ones undelivered through failure to notify Elsevier of change of address. Single copy and back volume information available upon request.

Please direct orders for this journal, changes of address, and claims for missing issues to Journal Fulfillment Department, Elsevier Science Publishing Co., Inc., 52 Vanderbilt Avenue, New York, NY 10017.

Please direct inquiries regarding the placing of advertisements in this Elsevier journal to Journals Advertising Coordinator, Elsevier Science Publishing Co., Inc., 52 Vanderbilt Avenue, New York, NY 10017. Telephone: (212) 867-9040.

Postmaster: send address changes to Controlled Clinical Trials, Elsevier Science Publishing Co., Inc., 52 Vanderbilt Avenue, New York, NY 10017.

© Elsevier Science Publishing Co., Inc., 1982. Application for second-class postage pending at New York, NY and at additional mailing offices. All rights reserved. This journal has been registered with the Copyright Clearance Center, Inc. Consent is given for copying of articles for personal or internal use of specific clients. This consent is given on the condition that the copier pay through the center the per copy fee stated in the code on the first page of each article for copying beyond that permitted by the U.S. Copyright Law. If no code appears on an article, the author has not given broad consent to copy and permission to copy must be obtained directly from the author. This consent does not extend to other kinds of copying, such as general distribution, resale, advertising, and promotional purposes, or for creating new collective works.

Proceedings of a Conference on Recent History of Randomized Clinical Trials Sponsored by the Kroc Foundation

Santa Ynez Valley, California
November 1981

Guest Co-Editors

Thomas C. Chalmers, M.D.
Mt. Sinai School of Medicine
New York, New York

Peter Amacher, Ph.D.
Santa Barbara, California

RECENT HISTORY OF RANDOMIZED CLINICAL TRIALS

The Kroc Foundation November 9-12, 1981

Front Row: Henry Buchwald, Janet Elashoff. Jeremiah Stamler, Helene Sancho-Garnier, James Holland, Christian Klimt.

Second Row: Thomas Chalmers, David Banta, Gordon Zubrod, Curtis Meinert. Thomas Parker.

Third Row: Sylvan Green, Edmund Gehan, Lloyd Fisher. Peter Amacher. James Schoenberger, Howard Rockette.

Controlled Clinical Trials

DESIGN AND METHODS

Official Journal of the Society for Clinical Trials

Volume 3, Number 3, 1982

SECTION I: SUPPORT OF CLINICAL TRIALS

THOMAS C. CHALMERS and PETER AMACHER
Introduction 163

CURTIS L. MEINERT
Funding for Clinical Trials 165

H. DAVID BANTA
RCTs and The Federal Government 173

SECTION II: CANCER TRIALS

C. GORDON ZUBROD
Clinical Trials in Cancer Patients: An Introduction 185

SYLVAN B. GREEN
Patient Heterogeneity and the Need for Randomized Clinical Trials 189

EDMUND A. GEHAN
Progress of Therapy in Acute Leukemia 1948–1981: Randomized versus Nonrandomized Clinical Trials 199

HOWARD E. ROCKETTE, CAROL K. REDMOND, BERNARD FISHER and PARTICIPATING NSABP INVESTIGATORS
Impact of Randomized Clinical Trials on Therapy of Primary Breast Cancer: The NSABP Overview 209

H. S. GARNIER, R. FLAMANT, and C. FOHANNO
Assessment of the Role of Randomized Clinical Trials in Establishing Treatment Policies 227

SECTION III: CARDIOVASCULAR DISEASE TRIALS

LLOYD D. FISHER and J. WARD KENNEDY
Randomized Surgical Clinical Trials for Treatment of Coronary Artery Disease 235

WILLIAM T. FRIEDEWALD and JAMES A. SCHOENBERGER
Overview of Recent Clinical and Methodological Advances from Clinical Trials of Cardiovascular Disease 259

HENRY BUCHWALD, LAURIE FITCH, and RICHARD B. MOORE
Overview of Randomized Clinical Trials of Lipid Intervention for Atherosclerotic Cardiovascular Disease 271

SECTION IV: GENERAL DISCUSSION SESSIONS

THOMAS C. CHALMERS
A Potpourri of RCT Topics 285

Edited by THOMAS C. CHALMERS and PETER AMACHER
Conference on Recent History of Randomized Clinical Trials:
Concluding Session—General Discussion 299

Introduction

This is the report of a conference sponsored by The Kroc Foundation held in November, 1981 at its headquarters in the Santa Ynez Valley of California. We, the organizers, thought the great number of randomized clinical trials since the mid 1960s and the attention given them in the establishment of new therapies made it advisable to take a broad look at them. Most meetings and publications on randomized clinical trials have dealt with specific trials or specific aspects of the methodology of trials. To keep the subject matter manageable, discussion was limited to the history of trials in cardiovascular disease and cancer.

Further, one of us (P.A.) began with the suspicion that those who designed and espoused randomized clinical trials had become dogmatic, a suspicion he found unwarranted. The articles here show that there has been continual innovation in the design of trials and that there is a healthy skepticism about some fundamental aspects of them.

One such aspect is the effect on the individual patient of being treated within a trial. The common assumption that care within a trial is on the average better than care outside of a trial was agreed to, and the usefulness of trials to patients who later receive better therapies established by trials is abundantly illustrated.

Some striking differences between trials in cardiovascular disease and in cancer emerged during the conference. There was also a lot of attention paid to the interactions between the scientific and ethical goals of trials, including the effects on both of differences in design and execution.

The papers published in this volume are edited versions of the oral presentations at the conference. Not all of the authors returned their edited manuscripts, and their formal presentations have been omitted. Discussion after each presentation has been either omitted, summarized by the editors, or included as corrected by the speakers.

The organizers wish to thank the Kroc Foundation and its president, Robert L. Kroc, for their support of the conference and this publication.

Thomas C. Chalmers
Peter Amacher

Controlled Clinical Trials 3:163 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030163-01$02.75

Funding for Clinical Trials

Curtis L. Meinert

The Johns Hopkins University, School of Hygiene and Public Health, Baltimore, Maryland

ABSTRACT: This paper provides a review of available funding sources for clinical trials, with emphasis on trials funded via the National Institutes of Health. Data from the NIH Inventory of Clinical Trials are used to summarize the cost and key design characteristics of NIH trials. The paper concludes with a discussion of some of the factors which add to the cost of trials.

KEY WORDS: *clinical trials funding, support, expenditures*

FUNDING FOR CLINICAL TRIALS

The sources of funds typically available for the support of clinical trials are listed in Table 1. Industry is a major source of support for trials that involve proprietary drug products. It is virtually impossible, however, to derive precise estimates of industry support for two reasons. First the cost for such trials, particularly Phase I and II trials, may not be distinguishable from other research and development costs within the budget of the companies and, secondly, companies are reluctant to provide cost figures even when available, because of the competitive nature of the drug field.

Some support is available through foundations and voluntary organizations such as the American Heart Association and American Cancer Society. However, the majority of funds available from these sources go for developmental work, including small-scale pilot studies, rather than for large-scale trials. The amount of support provided by academic institutions is impossible to assess since it comes primarily in the form of salary support for research personnel at those institutions.

You will note that third-party sources are listed as well. Here I am referring to insurance carriers such as Blue Cross/Blue Shield, Medicare/Medicaid, etc. Most, if not all, third-party payment sources for medical care have a proscription against payment for any procedure carried out solely for research purposes. However, procedures that are not only part of a study protocol but that can also be justified for routine patient care may be covered. My own view is that strict proscription is ill advised. Third-party payers must be encouraged to take a more positive approach toward support of clinical trials.

Received June 14, 1982; accepted July 1, 1982

Address reprint requests to: Curtis L. Meinert, Ph.D., The Johns Hopkins University, School of Hygiene & Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA.

Controlled Clinical Trials 3:165–171 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030165-07$02.75

Table 1. Potential Sources of Support for Clinical Trials

Government grants and contracts
Industry
Foundations and public funds
Academic research institutions
Third-party sources

It is quite conceivable that funds used by Medicare/Medicaid or Blue Cross/Blue Shield, if used for support of trials designed to evaluate specific care procedures, could ultimately help reduce the outlay for ineffective or unnecessary procedures.

The only detailed analysis I have been able to perform is for trials funded by the Veterans Administration (VA) and the National Institutes of Health (NIH). I say "only" because there are, to be sure, other agencies within the federal government that provide support for clinical trials but data on this is not readily available. VA expenditures for multicenter trials are recorded in Table 2 [1]. The amount of support provided in 1981 for multicenter trials represents about 7% of the total VA research and development (R & D) budget.

Table 3 provides expenditure data for NIH sponsored trials. These figures are based on data from the NIH Fact Book [2] and NIH Inventories of Clinical Trials for fiscal years 1975–79 [3]. NIH Inventories for 1980 and 1981 are not available.

The inventory listings are based on data supplied by the various Institutes of the NIH. A clinical trial, as defined by the NIH, is

> A scientific research activity undertaken to define prospectively the effect and value of prophylactic/diagnostic/therapeutic agents, devices, regimens, procedures, etc., applied to human subjects.

Table 2. VA Expenditures for Multicenter Clinical Trials, by Fiscal Year

Fiscal year	Total R & D budget[a]	Multicenter clinical trials[a]	Cost as percent of total budget
1970	58.1	1.8	3.1
1971	60.9	1.8	3.0
1972	69.1	1.8	2.6
1973	78.6	2.4	3.1
1974	81.8	4.3	5.3
1975	95.4	5.4	5.7
1976	101.6	5.9	5.8
1977	109.6	5.8	5.3
1978	118.0	6.3	5.3
1979	126.3	8.5	6.7
1980	137.7	9.0	6.5
1981	137.5	9.7	7.1

[a] Millions of dollars

Table 3. NIH Expenditures for Clinical Trials by Fiscal Year

Fiscal year	Total expenditures[a]	Number of trials	Expenditures[a] for clinical trials	Cost as percent of total
1975	2,093	755	87.8	4.2
1976	2,302	926	120.6	5.2
1977	2,544	746	105.3	4.1
1978	2,843	845	122.3	4.3
1979	3,190	986	136.2	4.3

[a]Millions of dollars

The definition is purposely broad to permit inclusion of the full spectrum of trials. It includes single-center as well as multicenter trials, extramural as well as intramural trials, randomized as well as unrandomized trials, and both controlled and uncontrolled trials.

The cost of clinical trials, relative to all NIH expenditures, has remained more or less constant over the time period covered in Table 3 and is, in any case, only a small fraction of the total NIH budget—4.3% of the 1979 budget. This is a fact worth noting in view of the concern expressed by some that the support provided for trials has reduced funding for more basic research. There is a symbiotic relationship between basic and applied research. One area of research cannot survive for long without the other. It is certainly reasonable to expect that Congress will have a more positive attitude toward the need for basic research if it is part of a larger program involving applied research as well.

Table 4 provides a breakdown of the 1979 cost figures by Institute, for all trials (NIH intramural as well as extramural) supported by the Heart (NHLBI), Cancer (NCI), and all other Institutes of the NIH. The median dollar cost per year of follow-up[1] is provided in the right-hand column of Table 4. The cost is highest for heart studies.

Tables 5–7 provide a summary of a few of the design features of single-

Table 4. 1979 NIH Expenditures for Trials by Institute

Institute	Number of trials supported	Expenditures[a]	Median per-patient per-year dollar cost
National Heart, Lung and Blood Institute	20	56.5	1657
National Cancer Institute	654	47.4	603
All Others	312	32.2	453
TOTAL	986	136.2	574

[a]In Millions of dollars

---

[1] The dollar cost per patient per year for a given trial is derived by dividing the total projected expenditures for that trial by the product of the anticipated number of patients to be enrolled in that trial and its projected length. The median dollar cost per patient per year for a given Institute is derived from the resulting figures for the individual trials sponsored by that Institute.

Table 5. Grant Supported Extramural NIH Trials

	Institute			
	NHLBI	NCI	Other	Total
No. of Extramural Trials				
Single Center	0	160	200	360
Multicenter	17	385	53	455
All Trials	17	545	253	815
No. Grant Supported				
Single Center	—	75	153	228
Multicenter	3	340	21	364
All Trials	3	415	174	592
Percent Grant Supported				
Single Center	—	47	76	63
Multicenter	18	88	40	80
All Trials	18	76	69	73

and multicenter trials in the 1979 Inventory by Institute. Of the 986 trials reported, 815 were classified as extramural (i.e., performed outside the NIH research facilities). Seventy-three percent of these were grant supported. The remainder of the 816 were contract supported or involved a mixture of grant and contract support (Table 5).

Nearly half (48%) of all trials included in the Inventory (extramural as well as intramural) were classified as multicenter. Seventy percent of the multicenter trials involved random assignment of patients to treatment, contrasted with just 50% of the single center trials. A similar, though not as striking a difference, is noted with regard to use of a control group. Sixty-five percent of the multicenter trials had a control group as compared to 55% of the single center trials (Table 6).

Multicenter trials, to no ones' surprise, are larger than single center trials. The median sample size (Table 7) for multicenter trials ($N = 166$) is close to three times that of single-center trials ($N = 60$). However, the per-patient

Table 6. Design Characteristics of 1979 NIH Trials

	Institute			
	NHLBI	NCI	Other	Total
No. of Trials				
Single Center	1	261	248	510
Multicenter	19	393	64	476
All Trials	20	654	312	986
Percent Randomized				
Single Center	100	45	56	50
Multicenter	100	67	75	70
All Trials	100	59	60	60
Percent with Control Group				
Single Center	100	38	73	55
Multicenter	100	59	94	65
All Trials	100	50	78	60

Table 7. Cost and Sample Size Data for NIH-Sponsored Trials, 1979 Inventory

Type of trial	Number of trials	Median sample size	Median per-patient per-year dollar cost
Single Center	510	60	$587
Multicenter	476	166	$523
All Trials	986	100	$574

per-year costs given in Table 7 are interesting in view of the oft-expressed opinion that "multicenter trials are too expensive." Their cost is actually less than that of single center trials ($523 vs. $587).

I pause here to note cost because it is an important issue, particularly in this age of shrinking research budgets. We must, as responsible members of the research community, consider ways to mount and carry out trials as inexpensively as possible. Mind you, I am not now addressing the question of the portion of the NIH pie that should be devoted to clinical trials but rather how one might stretch the dollars available for trials. The issue of what proportion of the budget should be devoted to trials is certainly an important one, but it is beyond the scope of my charge.

Both single and multicenter grant supported trials have a lower per-patient per-year cost than contract-supported trials although the difference is most striking in multicenter trials (Table 8)—due perhaps, to differences in the complexity of trials funded under the two mechanisms. However, at the same time, we should not lose sight of the difference between the two funding approaches. Grant-supported trials are generally investigator-initiated and are funded for fixed periods of time with ceiling levels set in advance of funding. Investigators are required to live within these bounds or must prepare a supplemental funding request in an attempt to obtain added support—a task not to be relished, even in the best of times. Contracts, on the other hand, are generally supported via a cost reimbursement mechanism that

Table 8. Patient Per Year Costs for NIH Extramural Grant vs. Contract Supported Trials

	Number of trials	Median sample size	Median per-patient per-year dollar cost
Single Center Trials			
Grant support only	228	60	385
Contract support only	127	100	448
Mixed support[a]	5	150	319
Multicenter Trials			
Grant support only	364	155	312
Contract support only	66	223	608
Mixed support[a]	26	650	334

[a] Both grant and contract, interagency agreements with grant or contract, intramural in conjunction with grant or contract.

allows an investigator to recover actual costs for justified expenses, even if they exceed initial estimates.

Table 9 lists a few factors that may contribute to increased costs. The remainder of this discussion relates to these items. There is no question that multiple aims complicate the design of a trial. It is natural to want to "kill two birds with one stone." In fact, it is efficient to do so, so long as the aims are complementary. There are, nevertheless, limits to the number of questions that can be answered in any one trial. The select nature of the study population means that it is of limited value for investigating the natural history of a disease process. To be sure, it may yield important clues in this regard, but they should be confirmed in a larger more general population, not subject to the selective forces of a trial.

Data processing costs are directly proportional to the amount of data collected. The tendency of clinicians to equate the data collection requirements of a study with those needed for patient care can lead to needlessly complicated data systems. The fact that an item of information is required for patient care is not reason enough to collect it for purposes of the trial. Data items that cannot be justified in relation to one of the main objectives of the trial should not be collected.

Generally, the larger the trial, the greater the urge to use it as a means of carrying out ancillary studies (i.e., studies that are carried out by a subgroup of investigators and that are not required to achieve the main aims of the trial). First and foremost, the studies should not interfere with the data collection requirements of the trial and, secondly, they should be compatible with its main objectives. Studies that are carried out in the context of a trial and that require funds from it should be approved via an outside peer review process before they are allowed to proceed. Ancillary studies should never be used as a means of maintaining investigator interest in the trial, but should have scientific merit in their own right.

The general axiom that haste makes waste is certainly true in clinical trials. Starting patient recruitment before the data forms have been adequately tested is always a mistake. The cost of mid-course corrections or of deficiencies in the data collected is nearly always greater than the costs associated with delays at the start of the trial due to development and testing of the data system.

The impetus to design more inexpensive trials must be pursued within reason and always with a recognition of the need for flexibility in the approaches used for funding, organizing, and managing them. We must never get to the point where, through our efforts to economize, we attempt to meet essential needs in a trial as if they can be purchased on a fee-for-service basis. There is already a tendency to move in this direction with the designation of so-called "set asides" in which the NIH is required to award monies for certain

Table 9. Factors That Add to the Expense of a Trial

Pursuit of secondary aims not related to the prime objective
Equating data collection needs for the trial with those needed for patient care
Ancillary studies
Start of patient recruitment without adequate preparation and testing of the data system

activities to small-for-profit businesses [4]. Cases in point are Request for Proposals (RFPs) released in recent years for data centers that can only be addressed by a small business [5]. Restrictions of this sort are not in the best interest of the field.

By the same token, undue emphasis on the cost of trials either by the sponsoring agency or by investigators can have long-term adverse effects on the climate in which they are carried out. Constant harrassment, whether by sponsoring agencies or study investigators, to achieve certain performance standards under stringent cost constraints will take all the joy out of an activity that, even under the best of circumstances, is tedious. Further, undue pressures can lead to corner cutting or, even worse, questionable data practices. Payments to clinics based solely on the number of patients seen are particularly suspect. Investigators who may be hard pressed, because of other budget cuts, may be tempted to look the other way as eligibility criteria are "reinterpreted" to help them meet their recruitment goal and enhance their funding at the same time.

A clinical trial, as suggested by Fredrickson [6], is an ordeal. As practitioners of trials, we must continue to pursue ways to design and carry them out so that they can be done efficiently and with relative ease, and so they are no more of an ordeal than necessary!

REFERENCES

1. Veterans Administration, Cooperative Studies Program, Medical Research Service, VA Central Office, Washington, D.C.
2. Basic Data Relating to the National Institutes of Health. U.S. Department of Health and Human Services, Public Health Service. Prepared by the Office of Program Planning and Evaluation and the Division of Research Grants. National Institutes of Health, 1981.
3. NIH Inventory of Clinical Trials: Fiscal Years 1975–1979. Research Analysis and Evaluation Branch, Division of Research Grants, National Institutes of Health, Bethesda, Md.
4. S.881: In the Senate of the United States. An Act: December 8, 1981, 97th Congress, 1st Session.
5. National Institutes of Health: Request for Proposal—Statistical support for cooperative groups engaged in intensive studies and investigations on cancer patients. National Cancer Institute, Bethesda, Md., RFP NCI-CM-87178-26, 1977.
6. Fredrickson DS: The field trial: Some thoughts on the indispensable ordeal. Bull NY Acad Med 44:985–993, 1968

RCTs and The Federal Government*

H. David Banta

ABSTRACT: As the federal government's role in health policy has grown, its involvement in the development, adoption, and use of medical technology has grown concomitantly. The government now makes a wide variety of decisions concerning technology. Many of these decisions involve randomized clinical trials (RCTs). The federal government probably funds or requires a majority of the RCTs carried out in the United States. The federal government also uses the results of RCTs increasingly as an important basis for its decisions, especially within the Medicare program. As the scientific basis for medical practice has improved through use of the RCT, policy decisions have been put on a firmer base.

KEY WORDS: *randomized clinical trial, government funding, funding for randomized trials*

RCTs AND THE FEDERAL GOVERNMENT

The federal government is obviously an important actor in the health policy arena in the United States. However, in the research area, the federal government largely determines the research agenda. In 1979, more than 61% of the entire societal investment in health research and development was made through the federal government [1]. The government's role has included funding research when the private sector has not perceived that it could profit from the results of the research. Thus, the government is the major supporter of basic biomedical research in this country. Evaluations of medical and surgical procedures by randomized clinical trial (RCT) is another type of research that one would not expect to find support easily. One would therefore expect that the federal government would be a major supporter for RCTs in this country, whether they were carried out in the government itself or by the private sector.

The results of RCTs are important to policy makers. Decisions concerning medical technology must be made continually. Such decisions include whether or not a drug should be marketed, whether or not a hospital will purchase a computed tomography (CT) scanner, and whether or not an insurance company will pay for a particular technologically-based service. Trial results themselves are of little utility to policy makers, who are not trained in their analysis

---

*Prepared for the conference The Recent History of Randomized Clinical Trials (RCTs), Sponsored by The Kroc Foundation, Santa Ynez, California, November 9–11, 1981.

Received June 14, 1958; accepted July 1, 1982.

Address reprint requests to: H. David Banta, M.D., M.P.H., Assistant Director, Office of Technology Assessment, Congress of the United States, Washington, D.C. 20510.

Controlled Clinical Trials 3:173–183 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030173-11$02.75

and interpretation. Thus, trial results must be synthesized into a form useful for policy making. This can either be done by a research program or by the program that needs the results.

The sections that follow describe (1) federal government support for RCTs, (2) federal government programs that synthesize results of research (and sometimes give advice based on the synthesis), (3) one particular program that requires RCTs as part of its decision-making process, the Food and Drug Administration's drug regulatory program, and (4) recent attempts to use RCTs and other research as the basis for policy decisions, especially those concerning reimbursement for services.

SUPPORT FOR RCTs IN THE FEDERAL GOVERNMENT

The principal biomedical research agency in the federal government is the National Institutes of Health (NIH). NIH currently expends approximately two-thirds of all federal dollars allocated to health research [2]. Although the legislation establishing NIH and its constituent institutes does not refer to efficacy or effectiveness, NIH considers that support for RCTs is implicit in its general mission of advancing knowledge and understanding of health, disease, and methods of preventing or curing disease. A number of other Federal agencies also support RCTs. These agencies include the Veterans Administration (VA), the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA), the Department of Defense, and various parts of the Department of Health and Human Services [3]. In 1975, for example, ADAMHA funded about $19.3 million in "treatment assessment research."

The NIH is by far the largest financial supporter of RCTs. In 1975, nine institutes of NIH expended $87.8 million in support of 755 clinical trials [4] (see Table 1). By 1979, the figure had risen to $136 million supporting 986

Table 1 National Institutes of Health 1975 Inventory of Clinical Trials (Number and Amount of NIH Support for Clinical Trials Active in Fiscal Year 1975 by Institute)[a]

Institute[b]	Number of trials	Amount of support
NIH	755	$87,817,682
NEI	20	$3,069,908
NHLBI	26	$42,698,181
NIAID	109	$3,090,915
NIAMDD	49	$3,365,641
NICHD	41	$3,856,325
NIDR	44	$1,717,514
NINCDS	59	$3,422,897
NIGMS	2	$55,500
NCI	405	$26,540,801

[a] Source: Reference 5.

[b] Note: NEI = National Eye Institute; NHLBI = National Heart, Lung, and Blood Institute; NIAID = National Institute of Allergy and Infectious Diseases; NIAMDD = National Institute of Arthritis, Metabolism, and Digestive Diseases; MICHD = National Institute of Child Health and Human Development; NIDR = National Institute of Dental Research; NINCDS = National Institute of Neurological and Communicative Disorders and Stroke; NIGMS = National Institute of General Medical Sciences; NCI = National Cancer Institute.

Table 2 National Institutes of Health 1979 Inventory of Clinical Trials (Number and Amount of NIH Support for Clinical Trials Active in Fiscal Year 1979 by Institute)[a]

Institute	Number of trials	Amount of support
NIH	986	$136,160,116
NEI	26	$8,605,609
NHLBI	20	$56,523,501
NIAID	120	$6,496,938
NIAMDD	67	$8,240,133
NICHD	32	$4,183,244
NIDR	26	$1,778,699
NINCDS	40	$2,660,949
NIGMS	1	$225,750
NCI	654	$47,445,293

[a] Source: Reference 6.

trials [5]. This increase more than keeps pace with inflation. As shown in Table 2, the cost per trial by institute varies widely, from almost $3 million for the National Heart, Lung, and Blood Institute (NHLBI) to about $54,000 for the National Institute of Allergy and Infectious Diseases (NIAID).

Not all these clinical trials are RCTs. Of the 755 trials funded in 1975, 373 were randomized and included a control group. One hundred fourteen of those were double blind in design [3].

As shown in Table 3, well over half of NIH-supported trials tested therapies. The more detailed 1975 data showed that 535 of the 755 trials tested drugs either in isolation or in combination with another type of technology [3]. More than 300 tested cancer chemotherapies. Only 25 treated surgical procedures, and few examined screening or early diagnosis.

The issue of testing for cancer chemotherapy is interesting. Although the NIH statute does not specify clinical trials, the NIH appropriation has specified testing of cancer chemotherapy since 1955. From 1956 to 1962, the annual amount of chemotherapy development and testing rose from $5 million to $43.7 million [6]. This earmarking, which has continued to the present day, has assured testing of cancer drugs. No other type of clinical trial has been specified in appropriations bills.

Probably the second largest support of clinical trials is the VA, which spends more than $100 million a year on health research [2]. Since the care in VA hospitals is without charge to either the veteran or the investigator, clinical trials can be done relatively cheaply within that system. Trials are done frequently, but the VA has no information on the numbers or types of trials overall. The VA generally funds trials of importance to its clients.

The VA has a Cooperative Studies Program that encourages the development of multicenter clinical trials. The total budget of this program increased from $2.8 million in 1972 to $9.6 million in 1979 [8]. This amount does not include investigator salaries or institutional overhead but does include expenses related to data collection and analysis. Studies often have non-VA funding, such as from NIH. During the period 1972–79, 57 studies were active. They were all prospective in design, and 49 included a control group selected by randomization. Well-known trials supported by the VA include RCTs of

Table 3 National Institutes of Health 1979 Inventory of Clinical Trials (Number and Amount of Support for NIH Supported Clinical Trials Active in Fiscal Year 1979 by Institute for Type of Intervention)[a]

	Total trials supported in FY 1979[b]		Therapeutic[b]		Type of intervention			
					Prophylactic[b]		Diagnostic[b]	
Institute	Number[c]	Amount[c]	Number	Amount	Number	Amount	Number	Amount
NIH	666	$112,847,367	494	$50,540,964	118	$58,875,778	53	$3,170,625
NEI	26	$8,605,609	22	$4,890,194	2	$3,415,997	2	$299,418
NHLBI	20	$56,523,501	10	$9,726,605	10	$46,796,896	—	—
NIAID	120	$6,496,938	57	$2,992,347	39	$2,697,064	24	$807,527
NIAMDD	67	$8,240,133	60	$7,680,072	4	$246,798	3	$313,263
NICHD	32	$4,183,244	16	$2,532,054	15	$1,629,175	1	$22,015
NIDR	26	$1,778,699	7	$779,051	17	$776,871	2	$222,777
NINCDS	40	$2,660,949	35	$1,565,020	2	$959,429	3	$136,500
NIGMS	1	$225,750	—	—	1	$225,750	—	—
NCI	334	$24,132,544	287	$20,375,621	28	$2,127,798	18	$1,369,125

[a] Source: Reference 6.

Note:[b] Trials in cooperative groups not included.

[c] One trial did not report amount of support. One trial did not specify type of intervention.

coronary bypass surgery, treatment for hypertension, and immune serum globulin for the prevention of post-transfusion hepatitis. Most VA cooperative trials deal with cardiovascular, psychiatric, and gastrointestinal disease [8]. The 24 trials ongoing in October 1981 include several dealing with coronary artery disease, dental treatment, and treatment of alcoholism and depression [9].

During the early 1970s, there was increasing concern about the benefits and risks of medical technology, as well as about the lack of data in specific cases [10,11]. This climate helped lead to an increase in funding of RCTs by NIH, as already noted. Four institutes of NIH nearly tripled their funding for clinical trials between 1971 and 1974 [12]. The National Heart, Lung, and Blood Institute, the only institute to have records over time, increased its funding more than 10 times over a ten-year period, as shown in Table 4. Nonetheless, the limited resources for clinical trials and the need for priority setting has led to the development of formal programs in the VA and NIH to plan trials [13,14]. For example, NHLBI has a decision process leading to initiation of a clinical trial and continuing through dissemination of results [15]. NIH has also established an interinstitute Clinical Trial Committee to deal with cross-institute issues.

Nonetheless, Congress became increasingly concerned about the lack of clinical trial results for policy making during the mid-1970s. There was no overall coordination of RCTs in the federal government, nor was there a mechanism to assure that important technologies were evaluated. The agencies mentioned above tended to evaluate technologies of interest to their program or clients. There was no mechanism to assure that technologies of importance to health planning efforts or reimbursement programs were studied. These findings led Congress to establish the National Center for Health Care Technology (NCHCT) by 1978 statute. Since its formation, the Center has made progress in coordinating activities concerned with the evaluation of medical technology. However, with a 1981 budget of about $4 million, NCHCT did not become a major research agency. It had power to determine the research priorities of other agencies. Its activities in synthesis and in giving coverage advice to the Health Care Financing Administration (HCFA) will be

Table 4 Obligations by the National Heart, Lung, and Blood Institute for Major Clinical Trials, 1963–74[a]

1963	$2,747,000
1964	$3,663,000
1965	$3,327,000
1966	$3,993,000
1967	$5,839,000
1968	$6,584,000
1969	$6,367,000
1970	$6,151,000
1971	$11,843,000
1972	$19,445,000
1973	$22,394,000
1974	$40,616,000

[a] Source: Reference 12.

described later. However, it should be noted that the Reagan budget allowed no funding for the Center, and it ceased to exist in December 1981, although some of the functions described later will continue in other loci.

SYNTHESIS OF RCT RESULTS IN THE FEDERAL GOVERNMENT

If RCTs are to be useful to practitioners and policy makers, their results must be generalized and put in relation to other knowledge. The process of carrying out this task is usually referred to as "synthesis." Synthesis requires the structured consideration of available knowledge. The root of the word is "putting together." It is important to note that synthesis is not the same as consensus. Consensus refers merely to agreement or to general opinion and need not be grounded in knowledge. There are complicated analytical techniques that can be used to assist synthesis, and several government programs carry out synthesis [16]. Synthesis activities tend to be in a different locus than research itself, perhaps because those involved in doing RCTs have limited interest in the application of the findings of their work [17].

The prominence of synthesis is relatively new. Traditionally, consensus has been the norm. Policy has typically been made by " . . . setting up expert committees—usually dominated by medical membership—to report on specific policy problems. Thus, the emphasis has been not so much on policy analysis as on achieving a policy consensus among those representing the very professionals that would have to implement any decisions made" [18]. Synthesis, on the other hand, seeks the best scientific knowledge. The reason that synthesis deserves discussion in an article on RCTs is that without good research results, synthesis results may have little validity. There are other reasons for synthesis. The decision maker is interested in aspects of a problem not covered in any one study. Multiple studies and sources of information may bear on a decision. Synthesis may bring the results of recent studies into prominence and put them in a form that can be more convincing than the raw data itself. Synthesis can also deal explicitly with methodological problems in various studies and may be able to correct for bias.

Probably the most important synthesis activity in the federal government is the Consensus Development Program of the NIH, which brings together concerned parties in order to seek agreement on the safety, efficacy, and appropriate conditions of use for various medical procedures and technologies. The method used by NIH is to select a broadly representative group of participants, including clinicians, epidemiologists, economists, experts in ethics, and laypeople. However, groups have tended to be dominated by clinicians, particularly those from academic medical specialties. Synthesis, although not specified in literature from the program, is a central part of the process. A thorough review of the scientific and clinical literature is usually done, and in many cases original data are collected. The task is done by members of the consensus group, with the assistance of NIH staff. The goal of each conference is to reach consensus among the panel [19]. The inability to reach consensus is usually taken to indicate that existing knowledge is inadequate. In preparing a consensus statement, the organizers report agreements and also identify areas where conclusions cannot be reached and areas in which further research is needed.

The Consensus Development Program was begun in 1977, and about 30 areas of medicine have been examined. The reports do not necessarily deal with technologies of policy importance. Reports themselves are voluminous, well-referenced, and written in professional jargon. Thus, they seem addressed primarily to the research and academic medical community.

Other parts of the federal government do synthesis. The NCHCT funded the preparation of a number of issue papers to synthesize information on a number of technologies of policy importance. NCHCT also held several national conferences, notably on alpha-fetoprotein and coronary bypass surgery, as part of a general activity that might be considered synthetic. The Office of Technology Assessment has completed a number of cases synthesizing information on specific technologies. Usually, these have been done to illustrate particular points and not to influence policy, but recent requests for studies from Congress seem more aimed at specific policy decisions [20]. Other organizations doing synthesis are less important, although there are several health policy centers in the private sector that perform synthesis, often with federal support.

FDA'S DRUG REGULATORY PROCESSES

The program most dependent on RCTs in the federal government is the premarketing drug regulatory program of the Food and Drug Administration (FDA). Although drugs have been regulated in the United States since 1902, it was the 1962 amendments to the Food, Drug, and Cosmetic Act that gave the FDA its power to require RCTs to demonstrate safety and efficacy.

Passage of the 1962 amendments was stimulated by the thalidomide tragedy of 1960–61 [21]. Secretary of Health, Education, and Welfare, Abraham Ribicoff, stated in hearings, "Under the existing safety clearance, we can consider effectiveness but not whether the claims of benefit are exaggerated . . . We have no basis for insisting that the manufacturer or distributor justify the full scope of each of the claims that he proposed to make" [22]. In 1961, a succession of physicians, testifying in favor of the amendments, stated that the ordinary doctor in clinical practice could not evaluate the efficacy of drugs. When the amendments were enacted into law, they provided that a drug should not be permitted on the market unless there was "substantial evidence" to support the claims made for it. Substantial evidence was defined as "adequate and well-controlled investigations, including clinical investigations . . ." FDA has interpreted the Act as requiring evidence from RCTs and has extensive regulations outlining the acceptable methods for doing such RCTs, including detailed design criteria. These RCTs are funded by industry but are often carried out in medical centers and by private physicians. The number of trials in progress at any one time is not known, but it certainly is in the hundreds.

The amendments also required a review of effectiveness for new drugs introduced between 1938 and 1962. Some 7000 drugs were in this status. Their efficacy was reviewed by a specific Committee of the National Academy of Sciences. The state of knowledge at that time is well illustrated by a statement of one of the participants in the scientific panels: "Many of the presentations submitted by manufacturers in support of the claims made for the use of their drugs consisted of bulky files of reports of uncontrolled observations and testimonial-type endorsements. The lack of substantial evidence based on well-controlled investigations by experienced investigators was conspicuous . . ." [23].

These amendments have assured that drugs on the market have been subjected to a minimal level of testing for efficacy and safety, in contrast to the situation for other types of technology. In addition, the benefits of the resulting information itself are considerable.

REIMBURSEMENT PROGRAMS AND RCTs

Reimbursement policy has two important relationships to RCTs. One is the extent to which insurance will pay for clinical testing such as RCTs. The other is the extent to which the results of RCTs are used in making reimbursement decisions.

Reimbursement for clinical testing is an important issue because there is little funding available for the explicit purpose of development of medical and surgical procedures, including the scientific assessment of their utility. Historically, patients have paid, largely through insurers, for this stage of technology development. For example, Roe has justified high fees for cardiac surgery as follows [24]:

> Historically, of course, there was justification for extraordinary fees in cardiac surgery. The developmental years of this field were indeed difficult, demanding innovative talent and an enormous amount of time . . . The early cardiac surgeon participated in the diagnostic studies and preoperative preparation, planned and directed the technical details . . .

Roe could have added that such evaluations as were done were generally done by the same surgeon, with the costs borne by insurance.

Until recently, it was relatively easy to obtain reimbursement for a new procedure. However, insurers are now tending to look for evidence of benefit before payment will be granted. This means that there is a strong incentive to request payment for a standard procedure, rather than a new one. For example, the new procedure, percutaneous transluminal coronary angioplasty (PTCA), is often billed as a coronary angiogram, an established procedure. Plasmapheresis is not covered routinely, since it is considered experimental in some states, but the fluids can be billed for, as can the care for the underlying condition [25].

This informal situation does not accommodate well to an entirely new procedure. Funding may not be available from any standard source, and this can lead to serious delays in availability. Indeed, a kind of "Catch-22" develops. The insurance companies will not pay for the procedure in the absence of evidence for its benefit, but there is no source of money for testing large enough to assure that all such procedures are tested. It has been proposed that insurance companies, and the Medicare program, develop an explicit policy of paying for experimental procedures when they are in the process of being evaluated [26]. Some insurers have already begun such a policy on a limited basis, but the Medicare program does not pay under such circumstance.

The other aspect of third-party payment is the use of data from RCTs and other evidence as the basis for coverage decisions. Traditionally, insurers have been passive, paying for whatever physicians do. Indeed, the Medicare Act specified that the legislation was not intended to influence the practice of medicine. However, cost concerns have led insurers to consider what they are paying for. The Medicare legislation included a statement that only expenses related to "reasonable and necessary care" would be reimbursed by the program.

Almost since its beginning, the Medicare program has had to make decisions about covering specific technologies. These have not been common, because the program has functioned in a very decentralized manner. However, with concerns about the costs and benefits of technology, Medicare has become more and more involved in specific decisions. Since the 1960s, advice on technical decisions has been sought from the Public Health Service, including NIH. By 1976, two people in the Office of the Assistant Secretary for Health carried out the function of coordinating advice. In 1977, as the function grew in importance and visibility, an Office of Health Practice Assessment was established in the Office of the Assistant Secretary for Health, one of whose purposes was providing assistance to the Medicare program. Finally, 1978 legislation specified that such advice would be provided by the National Center for Health Care Technology (NCHCT). With the disappearance of NCHCT, the future of this activity is uncertain.

Part of this task has already been described under consensus activities. However, the most important part is a staff function of gathering information on specific technologies and advising the Health Care Financing Administration as to whether it should or should not pay for a specific technology [26]. In March 1981, an Office of Technology Assessment staff paper [27] noted that NCHCT inherited a backlog of 58 requests for advice when it began operations and had collected 49 new issues. At that time, 35 issues were in the process of being examined, while the others had resulted in advice to HCFA. The most prominent recent example has been heart transplant. After much consideration, HCFA has decided not to pay for heart transplant pending results of a study to be funded by HCFA itself. Recent advice has included such issues as food allergy testing and plasmapheresis for rheumatoid arthritis [28]. Financial implications of this advice can be significant. The Center for the Analysis of Health Practices of Harvard University estimated that four decisions not to reimburse based on NCHCT advice could save the Medicare program from $88 to $959 million over a ten-year period. The four technologies examined were endothelial cell photography, dialysis for schizophrenia, hyperthermia for cancer, and radial keratotomy for myopia [29].

DISCUSSION

With concerns about the costs and benefits of medical technology, the RCT has assumed a new importance in federal government decisions. The idea has gradually developed that evaluations of technology, including RCTs, can play an important role in making policy decisions in the various areas of government activity. The government has been involved for some years in requiring RCTs of drugs as part of a regulatory process. More recently, the funding of RCTs has become an issue. Finally, the function of synthesis and giving advice on policy issues has become prominent during the past five years.

While much remains to be done, the approach to medical technology is much more systematic now than it was a few years ago. Pieces of a system to do studies for policy purposes, to synthesize their results, and to provide advice based on the studies and syntheses has gradually developed. While NCHCT itself no longer exists, the function of giving advice to Medicare has so far been explicitly preserved in the Office of the Assistant Secretary for Health. This is not to minimize the difficulties of developing a new emphasis for government programs. In particular, the American Medical Association and the Health Industry Manufacturers Association have sought the dissolution of NCHCT. Physicians have concerns, in part justified, about what they see as a move in the direction of the content of medical care being determined by bureaucrats and administrators. Manufacturers are concerned both about profits and about the effects of these new policies on medical innovation.

Despite resistance, the development of these new activities has been rapid. The visibility of the RCT as a tool for decision making has been growing. The RCT can only become more important as a basis for government health policy decisions [30].

REFERENCES

1. National Center for Health Statistics, National Center for Health Services Research: *Health United States. 1980*. DHHS publication no. (PHS) 81-1232. Hyattsville, MD: Department of Health and Human Services, 1980, p 295
2. National Center for Health Statistics, National Center for Health Services Research: *Health United States. 1980*. DHHS publication no. (PHS) 81-1232. Hyattsville, MD: Department of Health and Human Services, 1980, p 296
3. Office of Technology Assessment: *Assessing the Efficacy and Safety of Medical Technologies*. Publication no. OTA-H-75. Washington, D.C.: U.S. Government Printing Office, 1978
4. *NIH Inventory of Clinical Trials, Fiscal Year 1975*. Bethesda, MD: National Institutes of Health, 1975
5. Op. Cit.
6. Personal communication, Staff of the National Institutes of Health, 1981
7. Personal communication, Staff of the Veterans Administration, 1981
8. Henderson WG: Some operational aspects of the Veterans Administration Cooperative Studies Program from 1972 to 1979. Controlled Clin Trials 1:209–226, 1980
9. Veterans Administration: Status report, Cooperative Studies Program. October 1, 1981
10. Cochrane A: *Effectiveness and Efficiency*. Abingdon, England: Burgess & Son, 1972
11. Illich I: *Medical Nemesis: The Expropriation of Health*. Random House, New York: Pantheon Books, 1976

12. House of Representatives, Committee on Interstate and Foreign Commerce: *Investigation of the National Institutes of Health.* Washington, D.C.: U.S. Government Printing Office, 1976
13. James KE: A model for the development, conduct, and monitoring of multicenter clinical trials in the Veterans Administration. Controlled Clin Trials 1:193–207, 1980
14. Levy RI, Sondik EJ: Initiating large-scale clinical trials. Manuscript. National Heart, Lung, and Blood Institute, Bethesda, MD
15. Levy RI, Sondik EJ: Decision-making in planning large-scale comparative studies. Ann NY Acad Sci 304:441–457, 1978
16. Office of Technology Assessment: *Strategies for Medical Technology Assessment.* Publication no. OTA-H-181. Washington, D.C.: U.S. Government Printing Office, 1982
17. Young DA: Communicating the findings from clinical trials. Manuscript. VA Administrative Scholars Program, Veterans Administration, Washington, DC, 1981
18. Klein R: The rise and decline of policy analysis: the strange case of health policymaking in Britain. Policy Analysis 2:459–475, 1976
19. Perry S, Kalberer JT: The NIH Consensus-Development Program and the assessment of health technologies: the first two years. New Engl J Med 303:169–172, 1980
20. Office of Technology Assessment: *The Implications of Cost-Effectiveness Analysis of Medical Technology. Background Paper # 3: The Efficacy and Cost Effectiveness of Psychotherapy.* Publication no. OTA-BP-H-6. Washington, D.C.: U.S. Government Printing Office, 1980
21. Temin P: *Taking Your Medicine.* Harvard University Press, Cambridge, MA: 1980
22. Quoted by Goodrich WW: The responsibilities and problems of government. In: *Drugs in Our Society,* Talaley P, ed. Baltimore: The Johns Hopkins Press, 1964, pp 141–147
23. Division of Medical Sciences: *Drug Efficacy Study, A Report to the Commissioner of Food and Drugs.* Washington, D.C.: National Academy of Sciences, 1969
24. Roe BB: The UCR boondoggle: a death knell for private practice? New Engl J Med 305:41–45, 1981
25. Bunker JP, Fowles J: Assessment of medical technology strategies: coverage and reimbursement options for experimental medical and surgical procedures. Prepared for the Office of Technology Assessment, Washington, DC, 1981
26. Towery OB, Perry S: The scientific basis for coverage decisions by third-party payers. J Am Med Assoc 245:59–61, 1981
27. Office of Technology Assessment: The National Center for Health Care Technology and Medical Technology Assessment Policy. Staff Paper. Congress of the United States, Washington, DC, 1981.
28. Perry S: Minutes of the Meeting of the Department of Health and Human Services Technology Coordinating Committee, July 8, 1981.
29. Center for the Analysis of Health Practices: *Impact on Health Costs for NCHCT Recommendations for Nonreimbursement for Medical Procedures.* Boston, MA: Harvard School of Public Health, 1981
30. Banta HD, Behney CJ: Policy formulation and technology assessment. Milbank Mem Fund Q 59:445–479, 1981

Clinical Trials in Cancer Patients:

An Introduction

C. Gordon Zubrod

Comprehensive Cancer Center of the State of Florida, University of Miami School of Medicine, Miami

As chairman of today's session, I extend my thanks to the Kroc Foundation and to the panel members for making these discussions possible. The objectives of this meeting are to review the recent history of randomized clinical trials in cancer, to assess the advances in medical practice that have resulted, and to ask whether we are smart enough to distinguish between those trials that are fruitful and those that are not. Before calling up our first speaker, I wish to take the chairman's prerogative of 10 minutes to give an eyewitness view of some of the forces and individuals that led to the first randomized clinical trials in cancer. There were two streams of influence—infectious disease chemotherapy and comparative studies of analgesics and hypnotic drugs. The infectious disease component started with the wartime malaria studies. Although no randomized trials were undertaken, the renal physiologists, Shannon, Earle, Berliner, and Taggart, and the pharmacologists, Brodie and Udenfriend, designed highly quantitative trials of antimalarial drugs with special attention to pharmacokinetics and to the immunological response of the host [1]. Dr. Chalmers was part of this team and, of course, later went on to many studies in controlled trials. That remarkable man, E.K. Marshall, Jr., of Johns Hopkins, who played the key role in the clinical use of the sulfonamides, had a good deal to do with the malaria trials and after the war I joined him in studies of penicillin [2]. The group at Johns Hopkins made a number of contributions to controlled trials.

About this time the studies of streptomycin in pulmonary tuberculosis by the Medical Research Council were published and had a profound influence on the Johns Hopkins group [3]. We had formed a clinical trials committee—Mac Harvey, Marshall, Tommy Turner, Perrin Long, and Bill Cochrane—and had intense discussions of the ethics and methodology of trials. The first clinical study of penicillin in pneumococcal pneumonia at Johns Hopkins was not randomized [4]. The second trial [5] was our first effort at a randomized trial and these were continued over winters for six years. When Tumulty and

---

Received June 14, 1982; accepted July 1, 1982.

Address reprint requests to: C. Gordon Zubrod, M.D., Director, Comprehensive Cancer Center of the State of Florida, University of Miami School of Medicine, P. O. Box 016960 (D8-4), Miami, FL 33101, USA.

Controlled Clinical Trials 3:185–187 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030185-03$02.75

I went to St. Louis for a year (1953–54), the pneumonia trials were transported and Tom Frei, chief resident on the medical service, had his first introduction to randomized trials [6].

In the early 1950s, Dr. Lou Lasagna and Dr. Al Owens joined Marshall's department and became interested in the active metabolites of chloral hydrate. This led to a clinical trial of chloral hydrate and trichloroethanol, which introduced us all to the problems of subjective bias and methods of blind studies [7]. Lasagna went on to work extensively with Henry Beecher at Harvard on many similar problems.

In 1954, when I went to the NCI, Jim Holland was just leaving to go to Roswell Park, and we agreed to join forces in the study of acute leukemia in children. A protocol was developed and a comparative study started. When Tom Frei and Jay Freireich came to NCI in 1955, this effort expanded into the acute leukemia B group, which Dr. Holland headed for so many years. This group was responsible for the first randomized trial in leukemia, published in 1958 [8].

In 1954, Congress had created the CCNSC to stimulate work in chemotherapy of cancer. Dr. Ravdin was chairman of the clinical panel, and Dr. Holland and I were members. At the first meeting I drew attention to the methodologies of clinical trials in infectious disease and hypnotic drugs and passed around a reprint of Dr. Lasagna's on clinical trials [9]. I convinced Chalmers, Lasagna, and Owens to become interested in cancer so that much of the past came full circle. This led to the formation, in 1955, of the Eastern Solid Tumor group and the first randomized trial in solid tumors [10]. Other institutions became interested, a granting mechanism was developed and a protocol committee was formed. In addition to myself, this committee had as members Marv Schneiderman, Wayne Rundles, and the late Arnold Seligman. For 3–4 years it approved every protocol of the newly developing cooperative groups and had a profound influence on randomized trials in cancer.

REFERENCES

1. Shannon JA, Earle DP, Berliner RW, Taggart JV: Studies on the chemotherapy of the human malarias. I. Method for the quantitative assay of suppressive antimalarial action in vivax malaria. J Clin Invest 27:66–74, 1948 (Suppl)
2. Zubrod CG: Comparative efficiency of single and multiple dosage regimens of the penicillins: Bull Johns Hopkins Hosp 81:400–410, 1947
3. Medical Research Council: Streptomycin treatment of pulmonary tuberculosis. Brit Med J 2:769–783, 1948
4. Tumulty PA, Zubrod CG: Pneumococcal pneumonia treated with aqueous penicillin at twelve hour intervals. N Eng J Med 239:1033–1036, 1948
5. Austrian R, Mirick GS, Rogers DE, Sessoms SM, Tumulty PA, Vickers WH Jr, Zubrod CG: The efficacy of modified oral penicillin therapy of pneumococcal lobar pneumonia. Bull Johns Hopkins Hosp 88:264–269, 1951
6. Frei E III, Auner CR, Van Metre TE Jr, Zubrod CG: A comparison of the efficacy of tetracycline and penicillin in the treatment of pneumococcal pneumonia. New Eng J Med 252:173–176, 1955

7. Owens AH, Marshall EK Jr, Broun GO, Zubrod CG, Lasagna L: A comparative evaluation of the hypnotic potency of chloral hydrate and trichloroethanol. Bull Johns Hopkins Hosp 96:71–83, 1955
8. Frei E III, Holland JF, Schneiderman MA, Pinkel D, Selkirk C, Freireich EJ, Silver RT, Gold CL, Regelson W: A comparative study of two regimens of combination chemotherapy in acute leukemia. Blood 13:1126–1148, 1958
9. Lasagna L: The controlled clinical trial: Theory and practice. J Chron Dis 1:353–358, 1955
10. Zubrod CG, Schneiderman M, Frei E III, Brindley C, et al: Appraisal of methods for the study of chemotherapy of cancer in man: Comparative therapeutic trial of nitrogen mustard and thiophosphoramide. J Chronic Dis 11:7–33, 1960

Patient Heterogeneity and the Need for Randomized Clinical Trials*

Sylvan B. Green

Clinical and Diagnostic Trials Section, National Cancer Institute, Bethesda, Maryland

ABSTRACT: The hierarchy of strength of evidence concerning treatment efficacy ranges from individual anecdotal case reports, through studies with nonrandomized controls, to randomized clinical trials. The importance of randomized trials is their ability to convince the medical community. With randomization, bias (whether conscious or unconscious) is avoided, prognostic factors (both known and unknown) tend to be balanced across treatment groups, the validity of statistical tests of significance is guaranteed, and problems with time trends are avoided. A study of thyroid carcinoma demonstrates the importance of prognostic factors, and indicates how they can be identified using a Weibull model and used to form a prognostic index. A study of prostatic cancer illustrates how covariate information can be used to select optimal treatment for different patient subgroups in situations where there is no single best overall treatment. The key to successful clinical trials is a renewed effort in developing and selecting the treatments to be studied.

KEY WORDS: *clinical trials, optimal treatment, prognostic factors, randomization, survival analysis*

INTRODUCTION

Knowing that we hear, in a conference such as this, occasional disparaging references to the "religion" of randomized clinical trials, the organizers apparently decided that someone from the National Cancer Institute should be invited to preach the gospel. So I shall present the closing benediction.

The "gospel" of randomization is based on a pyramid (Table 1) conveying the hierarchy of strength of evidence concerning the efficacy of treatment. The spectrum of evidence ranges from individual anecdotal case reports to findings confirmed in multiple randomized trials. It is certainly true that efficacious treatments can be identified in the lower numbered steps of the pyramid. If an investigator happens to "hit a home run" in such a manner and discovers a cure for a previously untreatable disease, then it may not be necessary to initiate further studies. Smaller gains, however, require more stringent proof. I would argue that without steps 5 and 6 (i.e., randomized

---

*Presented at the Kroc Foundation Conference on Recent History of Randomized Clinical Trials, November 10, 1981.

Received June 14, 1982; accepted July 1, 1982.

Address reprint requests to Sylvan B. Green, M.D., National Cancer Institute, Landow Building, Room 5C-09, 7910 Woodmont Avenue, Bethesda, MD 20205.

Controlled Clinical Trials 3:189–198 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030189-10$02.75

Table 1 Hierarchy of Strength of Evidence Concerning Efficacy of Treatment

1. Anecdotal Case Reports
2. Series Without Controls
3. Series With Literature Controls
4. Series With Historical Controls
5. Single Randomized Clinical Trials
6. Confirmed Randomized Clinical Trials

trials), it is impossible truly to convince the scientific community about the efficacy of treatment.

This situation is especially true for studies of chronic diseases like cancer. There is a recurring theme that can be observed in conferences such as the present one. Whenever the participants are presented with results from nonrandomized studies, they cannot argue that the results are wrong—the results may very well be true—but the study fails to convince the audience. And if it fails to convince the people participating in the conference, it will fail to convince people in the medical community. Then, I would ask, what good has the trial been? Even if some money has been saved by not having a randomized trial, the principal objective has not been accomplished.

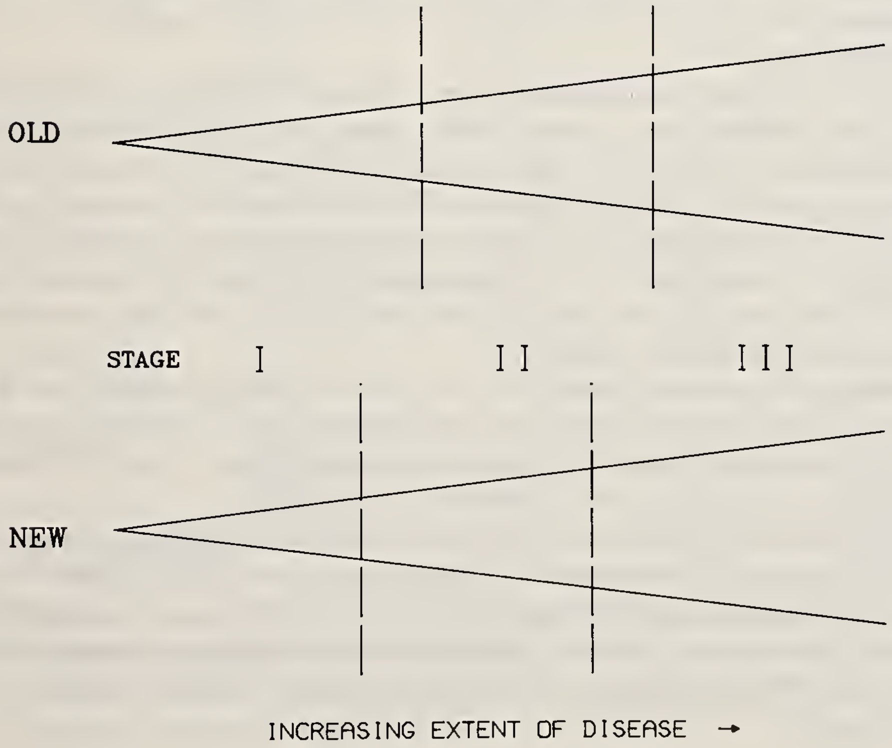
RANDOMIZED TRIALS VERSUS TRIALS WITH HISTORICAL CONTROLS

The question of whether or not a study is convincing is related to the issue of comparability of treatment groups. When groups are not comparable with respect to prognostic factors, adjusted analyses can be performed. However, difficulties may arise with historical controls. Given that we must adjust, we therefore must have the information needed to do the adjustment. We may know on which variables we want to adjust, but we may have trouble obtaining the data for the historical controls. This may not be a problem for a major institution that is in the business of conducting such studies, but sometimes the needed information just is not there. Often it is not possible to go back and obtain missing data when dealing with the past. Either a needed result was never obtained, or the determination was made but the data are now lost.

A different and even more difficult problem relates to the large number of unmeasured or unknown prognostic factors. For either of the diseases being discussed at this conference, whether it be cancer or cardiovascular disease, in spite of all the advances we have made in learning about the natural history of the disease, there is still so much we do not know. The imbalances in these factors are what can hurt us in nonrandomized studies, whereas randomization assures us that the groups will be alike on the average. Time trends are a very important aspect here. The patient population is changing, and these changes may be quite subtle. Changes in details of treatment and changes in supportive care may also introduce subtle and unconscious biases. Diagnostic methods are changing, and we are diagnosing diseases earlier, which makes survival look better. Survival is usually measured from the time of diagnosis. All one has to do is improve disease detection, and although nothing different has been done to affect the course of the disease, and although the patient is not living longer, it appears that there is improved survival.

An example of the subtle effect of time trends is illustrated in Figure 1. Suppose that the only difference between a historical group and a study group is that diagnostic methods have improved, and these methods are used to stage the disease into three categories (e.g., local, regional, and distant disease for cancer patients). I assume a simplified model in which the cancer progresses through these three stages, and the improved diagnostic techniques allow earlier detection of the spread of the disease. Examples of such techniques include new serum markers (and more sensitive assays), CAT scans, etc. In Figure 1, these changes mean that we detect the occurrence of regional spread (stage II) a little earlier, and we pick up metastatic disease (stage III) earlier, and therefore the dividing lines have shifted to the left. What is the result? Using stage II as an example, some of the patients who previously would have been diagnosed as stage I are now known to be stage II. These patients are not as sick as other stage II patients, so this addition of patients having better prognosis improves the survival of the stage II group as a whole. At the same time, some of the patients who previously would have been labeled stage II, the sickest ones, now are put in stage III. Thus, even if the new treatment is identical to the old one, it is going to look better for the present stage II group compared to the previous one. Likewise, stage I has had some of its sickest patients removed, so it would look better, and stage

Figure 1 Example of the effect of time trends in diagnostic procedures on composition and comparability of various stages of disease. Vertical spacing between the two slanted lines of each pair represents actual extent of disease.



III has been diluted by some patients who are not as sick as the others, so it will also look better.

To summarize this example, an investigator studies a new treatment, performs an analysis stratified on the important variable, stage of disease, and finds improved survival compared to historical controls in stage I, in stage II, and in stage III. Observing this result in all three categories, he is convinced that it is true. He thinks that obtaining the same result for all three stages means it could not be due to chance, and he believes something important has been found. In fact, nothing at all has been found. The old and new treatments are absolutely identical, and all that has occurred is an improvement in diagnostic techniques that has changed the labels given to the patients but not their prognosis. Such occurrences are sometimes very difficult to detect, but they certainly take place and can definitely lead to biased results. This is not a problem in a concurrently controlled, randomized trial.

The three major advantages of randomized trials which have been discussed in the literature [1,2] are (1) bias in assigning treatments is avoided, whether the bias be conscious or unconscious—and much of it is unconscious; (2) prognostic factors, both known and unknown, tend to be balanced across treatment groups; and (3) randomization guarantees validity of statistical tests of significance used to compare treatments. We have chosen our α-level, so that we know the type I error and can proceed with the analysis. If we observe that, even with randomization, an imbalance has occurred, we can still take account of prognostic factors in the analysis, perhaps using mathematical models, but because of randomization we are not required to do so. The randomization has protected the significance level, and the probability of imbalance is quantifiable. In the absence of this protection when nonrandomized controls have been used, one often must rely on mathematical assumptions, and the believability of the conclusion depends on the validity of these assumptions. When randomization has been used and imbalances have nonetheless occurred, the best advice is to present both the adjusted and unadjusted analyses.

With concurrent, randomized controls, time trends are not a problem since they affect all treatment groups in the same way. This is true for blocked randomization, but it indicates a possible problem with adaptive allocation schemes applied to chronic diseases. These schemes are interesting mathematically, and some people consider them to be attractive ethically. However, if there are time trends (e.g., in prognosis of the entering patients), one can get hurt by changing the ratio of allocation, since then the overall groups would no longer be comparable.

Missing data are less likely in a randomized trial because there is a prescribed protocol for data collection, and information is collected concurrently from all treatment groups. Thus one does not have to worry about what information was or was not available in the past.

For all of the above reasons, results from randomized trials are more likely to be convincing. I do not want to imply that medical data bases and other sources of data from nonrandomized studies cannot be used for learning about disease. For example, they are quite useful for studying prognostic factors. However, there are dangers in making treatment comparisons on the basis of such data.

THE INVESTIGATION OF PROGNOSTIC FACTORS

I shall use a study of thyroid carcinoma by the EORTC Thyroid Cancer Cooperative Group as an example both of the importance of prognostic factors and of how such data can be analyzed and applied. The data come from a registry of thyroid cancer patients, and thus this is also an example of using data from nonrandomized studies. The details of the study and analysis have appeared elsewhere [3]. Many other examples of prognostic factor analysis have appeared in the literature; an expository review of this subject has been written by Armitage and Gehan [4].

In the EORTC thyroid cancer study, a number of variables were investigated for prognostic importance, and after a multivariate analysis using a Weibull model of survival time, five variables were identified, each of which added to the ability to predict survival even in the presence of the other variables. One straightforward way to illustrate the prognostic effect of each of these variables individually is to examine the death rate for different categories (Table 2). The death rate is calculated under the assumption of exponentiality (rather than the more general Weibull assumption) and is defined here as deaths per 1000 patient-months of follow-up.

The coefficients from the Weibull model were used to form a prognostic index (Table 3). Because it was designed for use by clinicians, the index was

Table 2 Prognostic Variables and Death Rates for Thyroid Carcinoma

Variable	Category	Number of patients	Number of deaths	Total months follow-up[a]	Death rate[b]
Age	0–30	91	4	5077.5	0.79
	31–40	79	9	4176.5	2.15
	41–50	84	25	3746.0	6.67
	51–60	87	30	3666.5	8.18
	61–70	119	69	3637.5	18.97
	71+	47	38	933.5	40.71
Sex	Female	342	103	15428.0	6.68
	Male	165	72	5809.5	12.39
Cell type	Anaplastic	77	68	760.5	89.41
	MED or FLD[c]	121	45	4769.5	9.43
	All other	309	62	15707.5	3.95
T-Category[d]	T0,T1,T2	401	106	18242.5	5.81
	T3	106	69	2995.0	23.04
Metastatic sites	None	421	122	19041.5	6.41
	Single	78	46	2142.0	21.48
	Multiple	8	7	54.0	129.63
ALL PATIENTS		507	175	21237.5	8.24

[a] Because only completed months were recorded, 0.5 has been added to all follow-up times.

[b] Expressed in deaths per 1000 patient-months.

[c] Medullary or follicular less-differentiated, provided no associated anaplastic.

[d] Represents extent of the primary tumor.

Table 3 Suggested Prognostic Index for Thyroid Carcinoma

Age at diagnosis (in years)
+12 if male
+10 if principal cell type is medullary or follicular less-differentiated, provided there is no associated anaplastic
+45 if anaplastic
+10 if T-category is T3
+Choose one of the following:
  + 0 if no metastases
  +15 if there is a single site of distant metastases
  +30 if there are multiple distant metastatic sites

Total score	Risk group
<50	1
50–65	2
66–83	3
84–108	4
≥109	5

formed by dividing all coefficients by the smallest one and rounding to integers. Age was included in the model as a continuous variable; the others were categorical. Using the prognostic index, we can divide the patients into risk groups based on the predictions of the model. Earlier in this conference an example was presented in which patients had been ranked in order of risk, based on the clinician being asked simply to look at the values of the prognostic factors and guess what order to put them in. The mathematical model makes the procedure more objective.

The differences among the observed (actuarial) survival curves for the five risk groups (Fig. 2) are readily apparent, and in fact the spread of the curves appears to cover the plane of the survival plot. This indicates the marked importance of these prognostic variables in thyroid carcinoma. One can easily imagine that, in a nonrandomized comparison of patients selected from such a group of thyroid cancer patients, differences in survival due to treatment could be totally swamped by imbalances in these prognostic factors, unless one therapy represented a dramatic improvement over another.

It is for this reason that it is important to have the protection of randomization when making definitive comparisons of treatments. It is true that in this particular example I have five prognostic factors that have been identified, and I can adjust on them. But there are others that I do not know about. Furthermore, even with respect to those that have been identified, I am uncertain as to how well they have been measured, and I am relying on assumptions inherent in the particular survival model used. Comparing the observed survival curves with the curves predicted from the model (Fig. 2) suggests that the model fits reasonably well, but it is certainly not a perfect fit. Such considerations may introduce doubt as to the validity of nonrandomized treatment comparisons based on applying this model to this set of data. On the other hand, one might well use these prognostic variables for stratification in a future randomized trial. Finally, this risk group analysis may indicate for what subgroups of patients it would be best to design a

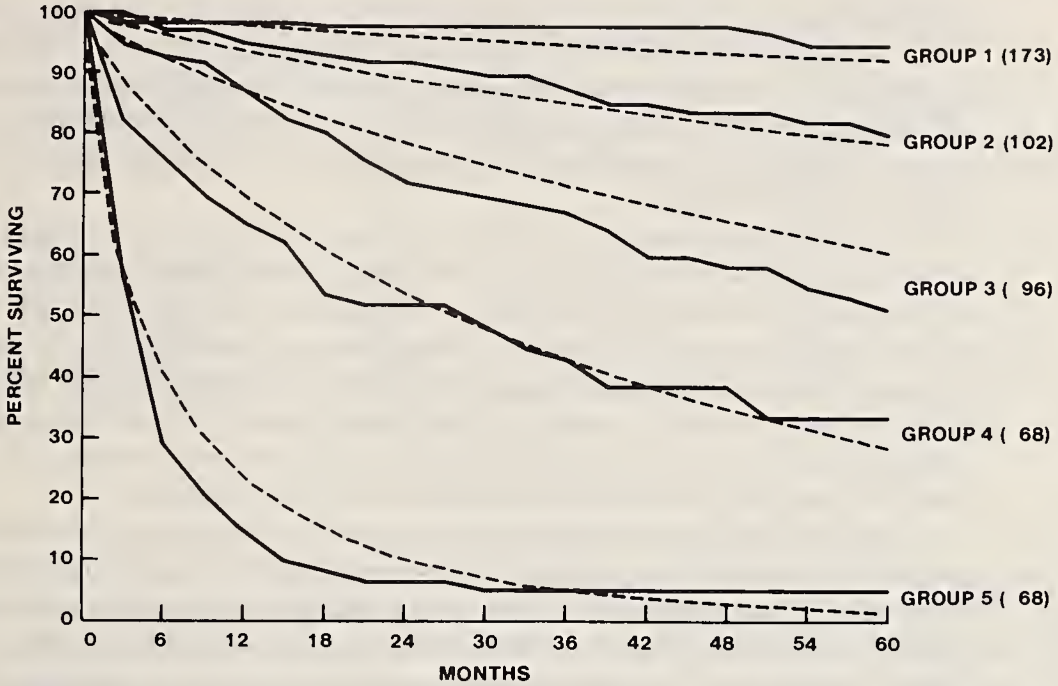


Figure 2 Observed (———) and predicted (-----) survival curves for the five risk groups of patients with thyroid carcinoma. Numbers of patients are shown in parentheses.

future trial. For the worst risk groups there is great room for improvement, and this might be a good place to concentrate efforts. Conversely, a very large sample size would be required to demonstrate improvements in survival for patients in the best risk group.

SELECTING OPTIMAL TREATMENT FOR PATIENTS BASED ON COVARIATE INFORMATION

The patient heterogeneity demonstrated by analyses of prognostic factors also suggests that all patients might not benefit from the same therapy, and all patients might not be primarily at risk for the same end point. The example I shall use involves determination of prognostic factors for a population of patients at high risk of death from both cancer and cardiovascular disease. The data are from a randomized trial performed by the Veterans Administration Cooperative Urological Research Group for stage III and stage IV prostatic cancer, comparing placebo and estrogen (diethylstilbestrol). This analysis has been reported in detail elsewhere [5].

One approach for determining whether treatment comparisons differ in various patient subgroups is to generate and compare survival curves for a number of different subgroups. The dangers of such an approach involve the multiple comparison problem and small sample sizes, with the resulting risk of spurious results. An alternative approach is to use a model, if one can be found that fits the data reasonably well, and to look for treatment–covariate interactions. The presence of a treatment–covariate interaction may imply (although not necessarily) that one treatment is better for one set of patients, and a second treatment is better for another set. In this situation the treatment of choice will vary depending on characteristics of the patient. Alternatively, an interaction may imply simply that one treatment is better (to a certain extent) for some patients and somewhat less so (or perhaps not different at all) for other patients.

An advantage of using a model is that one can perform a global test for all treatment–covariate interaction terms and, if that test is insignificant, stop at that point and avoid the multiple comparison problem. Alternatively, one can continue to look for specific interactions, but perhaps with a decreased degree of belief. However, if the global test for interactions is significant, one may definitely want to investigate individual terms. This is still exploratory data analysis, and one should not believe the nominal p values, but at the least the analysis provides hypotheses to test in the future.

In the prostatic cancer example, six prognostic variables and treatment were needed in an exponential model of survival time, as were two treatment interactions, one with tumor stage-grade and one with patient age (Table 4). This model implied that there was a different optimal treatment for different patient subgroups. If a patient had a tumor with a high stage-grade (i.e., an advanced tumor), then the treatment of choice was estrogen. On the other hand, if a patient was elderly, and especially if the patient had a tumor with a low stage-grade, the patient would actually have done better on placebo. One might postulate that this was because the risk of cardiovascular death would have been greater than that of cancer-related death. The values of age and stage-grade for any particular patient would determine which terms in the model predominate.

The model was actually fit considering all causes of death together, without identifying specific causes of death. To help validate the conclusions, it was decided to divide the patients into a group whose optimal treatment (according to predictions of the model) would have been placebo, and a group whose

Table 4 Prognostic Variables[a] and Regression Coefficients from Exponential Model for Prostatic Cancer

Variable	Regression coefficient[b]	Standardized coefficient[c]	Two-tailed p value
Intercept term	8.790	4.75	<.0001
Hemoglobin	8.337	2.71	.0067
Performance status	11.692	1.98	.0477
History of cardiovascular disease	9.863	4.48	<.0001
Tumor stage-grade category	13.433	3.92	<.0001
Patient age at diagnosis	0.804	0.28	.7794
Size of primary lesion	17.820	3.19	.0014
RX[d]	−3.105	−1.37	.1707
RX * stage-grade interaction	−8.447	−2.00	.0455
RX * age interaction	9.975	2.45	.0143

[a] All variables were coded as two-level or three-level categorical variables.

[b] Expressed in deaths per 1000 patient-months.

[c] Equal to the regression coefficient divided by its standard error.

[d] RX represents treatment group; placebo was coded as 0 and estrogen was coded as 1.

Table 5 Cause of Death by Actual and Optimal Treatment for Prostatic Cancer Patients

Optimal treatment	Placebo		Estrogen	
Actual treatment	Placebo	Estrogen	Placebo	Estrogen
Alive	11	10	46	72
Dead				
Prostatic cancer	5	5	72	43
Cardiovascular	16	26	45	52
Other causes	18	14	28	20
ALL PATIENTS	50	55	191	187

optimal treatment would have been estrogen, and then compare those groups by actual treatment received and actual cause of death (Table 5). Since the patients had been randomized at the beginning of the study, each patient had an equal chance of receiving either placebo or estrogen. Therefore, regardless of the category of optimal treatment to which a patient is assigned by the model, there is a 50% chance that the actual treatment would have corresponded to the optimal one and a 50% chance that it would have differed. Such balance was approximately realized, as can be seen in the bottom line of Table 5.

Among those patients whose predicted optimal treatment was placebo, those who actually had received estrogen experienced an excess of cardiovascular deaths, whereas the prostatic cancer deaths were balanced (and small in number). Conversely, among those whom the model predicted would do better on estrogen, the group that actually had received placebo had a much higher number of prostatic cancer deaths. There was a slight excess of cardiovascular deaths among those receiving estrogen, but the difference was small; clearly the estrogen was more beneficial than detrimental. This example illustrates the concept of treatment of choice and shows that there might not be a single best overall treatment in a clinical trial.

CHOOSING TREATMENTS FOR CLINICAL TRIALS

I now briefly mention one preliminary result from an analysis on which I have been collaborating with Dr. Maurice Staquet of the EORTC in Brussels. He has reviewed the medical literature, as well as NCI and EORTC files, to identify single drugs, excluding hormones, that had been clinically tested for effectiveness against solid tumors of any of four sites: breast, head and neck, ovary, and lung (small cell carcinoma). The definition of "clinically tested" (i.e., evaluable) required at least one study, randomized or not, with 14 or more patients. The definition of "active" (i.e., clinically effective) required responses, whether complete or partial, in at least 10% of the treated patients.

These definitions are admittedly somewhat arbitrary, but some useful results emerge (Table 6). Among thousands of drugs that have been tested preclinically in various animal systems, 69 were identified as having been

Table 6 Clinical Effectiveness of Anticancer Drugs for Four Solid Tumor Sites

Site	No. clinically tested[a]	No. active[b]
Breast	61	21
Head & Neck	35	14
Ovary	23	13
Lung (Small cell carcinoma)	31	16
Any of the four sites	69	36

[a] Requires at least one study with 14 or more patients.

[b] Requires responses in at least 10% of treated patients.

clinically tested, and 36 of these were active according to the above definition of minimal activity. Only 8 or so of these drugs met more stringent criteria limited to complete responses. There are a number of explanations for failure of agents to be tested clinically: some were found to have unacceptable toxicity, some could not be formulated satisfactorily, and many gave negative or unimpressive results in the animal studies. Perhaps the major lesson here, however, is that clinical trials are only as good as the treatments they are designed to study. Perhaps what we really need are increased efforts in the development of new therapies and better ways to predict from preclinical studies which of these therapies should have the highest priorities for inclusion in the clinical trials we design.

ACKNOWLEDGMENTS
The author thanks Julie Paolella for the typing of the manuscript.

REFERENCES

1. Byar DP, Simon RM, Friedewald WT, Schlesselman JJ, DeMets DL, Ellenberg JH, Gail MH, Ware JH: Randomized clinical trials. Perspectives on some recent ideas. New Engl J Med 295:74–80, 1976
2. Green SB: Randomized clinical trials: Design and analysis. Seminars in Oncology 8:417–423, 1981
3. Byar DP, Green SB, Dor P, Williams ED, Colon J, van Gilse HA, Mayer M, Sylvester RJ, VanGlabbeke M: A prognostic index for thyroid carcinoma. A study of the EORTC Thyroid Cancer Cooperative Group. Europ J Cancer 15:1033–1041, 1979
4. Armitage P, Gehan EA: Statistical methods for the identification and use of prognostic factors. Int J Cancer 13:16–36, 1974
5. Byar DP, Green SB: The choice of treatment for cancer patients based on covariate information: Application to prostate cancer. Bull Cancer (Paris) 67:477–490, 1980

Progress of Therapy in Acute Leukemia 1948–1981:

Randomized versus Nonrandomized Clinical Trials

Edmund A. Gehan

The University of Texas System Cancer Center, M.D. Anderson Hospital and Tumor Institute, Houston, Texas

ABSTRACT: This paper gives a review of selected clinical studies in acute leukemia over the last 33 years and addresses two questions: What advances in the practice of medicine have resulted from clinical trials, both randomized and nonrandomized? What generalizations can be made about clinical trials which have led to advances in treatment? The review demonstrates that nonrandomized studies have been the primary mechanism for establishing the efficacy of new therapies. Such studies have also been utilized to determine patient characteristics related to prognosis and some important concepts of treatment (e.g., independent action of some drugs). Randomized studies have tended to yield positive results concerning treatment differences when there were prior suggestions of activity from nonrandomized studies. Analyses of randomized studies have been made to determine patient features related to prognosis and concepts for the treatment of patients (e.g., independent action of some drugs and the importance of treating patients in a disease-free state).

KEY WORDS: *review of clinical trials in acute leukemia, randomized vs. nonrandomized studies*

INTRODUCTION

The purpose of this article is to review some of the major accomplishments of clinical trials in acute leukemia, both randomized and nonrandomized, over the last 33 years. Prior to 1948, there were no effective treatments for acute leukemia, and patients died in a median time of approximately nine weeks. Currently, there are very effective treatments for inducing and maintaining complete remission, both in children and in adults. The complete remission rate is approximately 60% for adult patients and is in excess of 80% for pediatric patients. A substantial percentage of both pediatric and adult patients currently being treated may even be cured of their disease. Consequently, the period 1948–81 has been one of significant progress in the de-

---

Received June 14, 1982; accepted July 1, 1982.

Address reprint requests to: Edmund A. Gehan, Ph.D., Professor of Biometrics, The University of Texas System Cancer Center, M.D. Anderson Hospital and Tumor Institute, Houston, TX 77030, USA.

Controlled Clinical Trials 3:199–207 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030199-09$02.75

velopment of therapies, in understanding the characteristics of patients related to prognosis, and in the development of new concepts of treatment.

It is worthwhile reviewing the progress of therapy in acute leukemia and addressing two questions: first, what advances in the practice of medicine have resulted from clinical trials, both randomized and nonrandomized? Practice of medicine is defined as that which is advocated by experts in the field, no matter how often such medicine is actually practiced. Second, what generalizations can be made about clinical trials that have led to advances in treatment and those that have not? The overall objective is to put in perspective the accomplishments that have resulted from randomized and nonrandomized clinical trials.

Review must necessarily be selective in the clinical trials that are considered. Because of the hundreds of clinical trials in acute leukemia that have been conducted over the last 33 years, this review cannot claim to be comprehensive. Rather, it is a review of clinical trials about which I am knowledgeable and which were at the forefront of the progress being made.

REVIEW OF CLINICAL TRIALS

The first demonstration that drug therapy was effective treatment in acute leukemia was by Farber [1] in a nonrandomized study of methotrexate in pediatric acute leukemia. For the first time, complete remissions of the disease were achieved and there was some hope for patients. Prednisone was evaluated by Pearson [2] and Farber [3] in 1950 and Burchenal [4] was the first (in 1953) to report the successful use of 6-mercaptopurine (6-MP). Burchenal reported a complete remission rate of 33% in 45 children with acute leukemia and a complete remission rate of 22% in 18 adults; both of these percentages are not very different from those obtained after a much larger number of patients had been studied.

Freireich et al. [5] reported a clinical trial of 6-azauracil in acute leukemia that had both randomized and nonrandomized elements. There was no effective treatment for patients who had already received what then were the only active chemotherapeutic agents (folic acid antagonists, adrenocorticosteroids, and purine analogs). Hence, new agents were studied in these patients, and no data were available concerning the natural history of this phase of the disease. In the randomized comparative study, there were 28 patients, 15 of whom received 6-azauracil and 13 of whom received placebo, whereas in the nonrandomized portion of the study there were 23 patients. There was objective improvement in approximately 30% of the patients in both the randomized and nonrandomized studies, with no evidence of a difference between 6-azauracil and control patients. Survival curves for each of the three groups of patients indicated a median survival time of approximately three months from onset of study to death. In the conclusions of the paper, 6-azauracil was found to have "no significant anti-leukemia activity": however, this study did contribute to defining the natural history of the advanced refractory phase of acute leukemia. Further, there was evidence that the nonrandomized study group acted in a similar manner to the comparative study group. This was a forerunner to some later cooperative studies in which there was no double-blind randomized comparison group but only a nonrandomized group.

Freireich and Gehan et al. [6] analyzed 178 consecutive patients who had an established diagnosis of acute leukemia and were treated between 1953 and 1958 at the National Cancer Institute. This paper established certain patient characteristics as related to prognosis and developed concepts of treatment that have proven useful subsequently. The percentage of patients having complete remissions was shown to decrease both by age group up to age 20 and subsequent to age 20. It was demonstrated that the survival of patients with acute leukemia can be attributed almost entirely to the period of time that patients spend with "hematologic improvement." By subtracting the periods of time that patients spent with "hematologic improvement" it was stated that "the risk of death during periods of active disease has not changed during the past ten years and the improved survival can be attributed directly to the duration of hematologic responses resulting from chemotherapy."

Another important concept was that the initial response of a patient was independent of the number of subsequent responses. At that time, some clinical investigators thought that patients could be classified as "responders" or "nonresponders" and the category of the patient predicted subsequent response and survival. The average number of later responses for patients having a first response of complete remission was 1.0, as was the average for patients who initially did not respond. Patients who had improvements or partial remissions initially had an average of 0.7 subsequent responses. The paper concluded "Response to subsequent courses of therapy was found to be independent of the response to the first course of therapy." From here it required only one further step to conclude that combining chemotherapeutic agents into one treatment regimen could increase the chance of an individual patient responding by the independent action of the drugs in the combination.

In 1961, Frei et al. [7] reported a randomized study that, in phase I, compared 6-MP vs. methotrexate vs. the combination of 6-MP and methotrexate. In phase II, patients failing after 6-MP received methotrexate and those failing after methotrexate received 6-MP. A total of 462 patients were entered into this study, and there were no overall differences in the survival experiences of the patients on the three treatment programs, the median survival time being 9–10 months for patients on each regimen of treatment.

An important part of this study involved the further development of the concept of independent action of drugs. In phase I, 26% of the patients had complete remissions after 6-MP treatment, 22% had complete remissions after methotrexate, and 42% of the patients responded to the combination treatment. As stated in the paper, "If it is assumed that when combination therapy is given, 6-MP and methotrexate act independently, the expected number of remissions for the combination may be calculated. For example, if 26% of children with ALL have complete remissions on 6-MP (the 6-MP CR rate) and in addition 22% (the MTX CR rate) of the 6-MP failures respond to methotrexate, the combined rate would then be 42%. This combined rate, when applied to the 40 children actually receiving the combination treatment, results in 16.9 as a predicted number of complete remissions, and this is very close to the 17 observed complete remissions. Thus, there is evidence in terms of remission induction that combined 6-MP and methotrexate provided a therapeutic advantage over either drug use alone. The data suggest that the advantage derives from the two drugs acting independently on the patient" [7].

Further, this study demonstrated that response in phase II of the study was almost the same as in phase I for those patients receiving single agents in each phase. The paper concluded: "Responsiveness to the second course of antimetabolite therapy was as good as that to the first course of treatment. This was true for remission rate, remission duration, and even for survival when appropriate corrections were made. Thus, responsiveness to drug therapy is maintained as the disease progresses temporally" [7]. This is an example of a randomized comparative study that confirmed and extended the concept of independent action of drugs and was a forerunner to subsequent studies involving multidrug regimens for patients.

In 1963, Freireich et al. [8] reported a prospective, randomized, double-blind, placebo-controlled, sequential study of 6-MP vs. placebo in the maintenance of remissions in pediatric acute leukemia. In the design of the study, patients were put into complete remission (in phase I) with steroid treatment and were then randomized to 6-MP or placebo (in phase II) which was administered double-blind. Patients were paired at each institution and the main end point was the length of complete remission between the two groups of patients in phase II. In phase III, patients relapsing after placebo treatment received 6-MP therapy. This study was analyzed several ways, each of which established that 6-MP led to substantially longer remissions than placebo. Among the 21 pairs of patients for whom a preference for 6-MP or placebo could be recorded, there were 18 preferences for 6-MP compared with 3 preferences for placebo, a highly statistically significant difference. The median length of 6-MP remission was 33 weeks compared with 9 weeks for those maintained on placebo. It is of interest that the difference between 6-MP and placebo treatment was also established in the nonrandomized portion of the study. Those patients relapsing on placebo who were later treated with 6-MP had remissions that were on the average 4 weeks longer than those on placebo. This study was the first to establish a very important concept, namely that patients with acute leukemia should be treated in the remission phase of the disease. It was a forerunner to adjuvant studies in other forms of cancer, such as breast cancer, in which treatments are administered when the patients are in a disease-free state. Further, it showed that a patient receiving placebo could be used as his or her own control for evaluating subsequent treatments. Such a study would not now be ethical or feasible, however, because of the responses that can be achieved with other therapies.

The first study in acute leukemia with a combination of four agents to induce remission was a nonrandomized study in 1964 of VAMP treatment by Freireich and Frei [9]. This study had a complete remission rate of over 60% and clearly established that multiple treatments could be combined to yield a higher response rate. An essential aspect was that each of the agents should have different side effects.

In the 1960s and the early 1970s, several new agents were recognized as being efficacious in nonrandomized single agent studies. Karon [10] showed that the vinca alkaloids could induce complete remissions. Freireich et al. [11]

established the value of cytosine arabinoside for the treatment of patients with acute granulocytic leukemia. In the early 1970s, French and Italian clinicians showed that adriamycin and daunorubicin were effective in inducing complete remissions.

The first publication demonstrating that the protected environment was useful for treating patients with acute leukemia was that of Bodey et al. [12] in 1971. This study was confirmed subsequently by Rodriguez [13] and some other investigators. In Bodey's nonrandomized study, there was no formal design; a total of 33 patients with acute leukemia received chemotherapy while in a protected environment (PE) and received prophylactic antibiotic regimens. No plan was made for a control group, but two control groups of 33 patients were selected from patients treated at about the same time to be as comparable as possible to the PE group with respect to chemotherapy regimen, diagnosis, age, sex, infection status, WBC, platelet count, time from diagnosis to therapy, and previous therapy. For each PE patient, a list of at least five potential pairmates was made, and the patients were ranked by Dr. Bodey according to his judgment of the probability of response. Control patients were selected to have equivalent or superior ranks to those for PE patients. To test the process of selecting control patients, a comparison was made between the two control groups of 33 patients; no differences were demonstrated in a paired analysis between these groups in complete remission rate, incidence of infection, or survival. Finding no difference between control groups, selected as indicated, provided a justification for comparing the PE vs. the control group. When PE patients were compared with the combined control groups, there was a suggestion of a difference in complete remission rate (61% for PE vs. 49% for combined control), but the main result was that there were significantly fewer episodes of both local and severe infection as related to neutrophil count in PE patients. Further, there was a significant improvement in survival time (median of 40 weeks for PE patients compared to 22 weeks for the combined control patients).

The paper concluded: "The use of nonrandomized patients as a control group has been considered unacceptable by some investigators. However, this approach was both necessary and advantageous for our study. Randomizing patients to either the PE group or control group would have resulted in the PE unit being unoccupied for substantial periods of time and would have required a much longer time to complete the study. Randomization would not have guaranteed that patients in the PE group and control group would be comparable, whereas our selection procedure ensured that the pairmates were as comparable as possible with respect to nine patient characteristics which have been demonstrated to affect prognosis" [12]. This paper was early evidence in cancer clinical trials that a nonrandomized study could lead to conclusions about treatments.

In 1978, Rodriguez et al. [13] reported a randomized study of a protected environment–prophylactic antibiotic (PE-PA) program. The design of the study was in the form of a two-by-two factorial where the factors were PE status (in vs. out) and antibiotic status [systemic (SA) vs. prophylactic (PA)]. The study demonstrated a significant improvement in complete remission rate for PE patients vs. non-PE patients (71% vs. 43%, $p < 0.01$), a significant advantage in survival time for PE patients (median of 80 weeks for PE, 50 weeks for non-PE, $p < 0.01$), and the incidence of infectious episodes according to neutrophil count was also significantly lower.

An example of a randomized, but ultimately inconclusive study was reported by Coltman et al. [14] in 1978. There were a total of 274 adult patients with acute leukemia who were assigned to one of three regimens: vincristine, prednisone, cytarabine (OAP); the same treatment plus 100 mg/m^2 of cyclophosphamide (COAP); or the same treatment plus 100 mg/m^2/day of daunarubicin (DOAP). In the most common group of patients, those with acute myelocytic leukemia (AML) not having received prior treatment, the complete remission rates were 37% for COAP, 35% for DOAP, and 43% for OAP. The differences among the complete remission rates were not near statistical significance, nor was there any evidence of a difference in length of survival by induction treatment ($p > 0.36$). Hence, the conclusion in the paper was essentially negative with respect to whether COAP or DOAP was better treatment than OAP.

However, an analysis had been made of patients treated by the group in the past and 6 prognostic factors were identified as related to the probability of complete remission. These were type of acute leukemia, age, hemoglobin value, logarithm of platelet count, weeks from symptoms to diagnosis, and logarithm of white blood count. Using the logistic regression equation, predicted probability of complete remission could be calculated for each patient based on prognostic features. At the three largest institutions, which entered 90 evaluable patients, the observed complete remission rate was 56%, compared to a predicted rate of 53% based on prognostic features. At all other institutions, which entered 128 evaluable patients, the observed complete remission rate was 34% compared with a predicted rate of 44%. At the three largest institutions, the difference between observed and predicted rates of complete remission was not near statistical significance ($p > 0.10$); however, the difference was significant at the other institutions ($p = 0.02$). When individual treatment groups were considered within the three largest and all other institutions, the only treatment that yielded results significantly different than predicted was DOAP treatment; the observed complete remission rate in 28 patients at the largest three institutions was 71%, compared with a predicted rate of 53% ($p = 0.03$). At all other institutions, the complete remission rate in 37 patients receiving DOAP treatment was 27%, compared with a predicted rate of 39% ($p = 0.10$). Results in these subgroups of patients were not reported in the published paper; however, the Southwest Oncology Group and other cooperative study groups have conducted other studies in acute leukemia using variants of DOAP treatment. The remission induction treatment in the Southwest Oncology Group and some other clinical groups consists of adriamycin combined with OAP treatment. Analyses of various prognostic factors were made to try and explain the differences in observed versus predicted response rates at the larger and smaller institutions; no particular factor could be identified explaining the differences. This is an example of a randomized prospective study, in which the results over all institutions indicated no evidence of treatment differences, but in which the evaluation of selected subgroups of patients nevertheless provided a lead for further studies.

In the last several years, statistical regression models have often been utilized to relate end point to patient characteristics and treatment. When complete remission is the end point, logistic models have been used. When length of complete remission or survival time is the end point, Cox's regression model has been used. A clinical trial was started in March, 1980, at the University of Texas System Cancer Center in which all previously treated patients with newly diagnosed acute leukemia have been entered into study. This is the first study in which the treatment of the patients is decided by calculating a predicted probability of complete remission utilizing a logistic regression model based on prognostic factors. This study, which has been described in Gehan et al. [15] and which is still in progress, is based on the following strategy: Conventional remission induction and maintenance therapy (AD-OAP for remission induction followed by various combination therapies for remission maintenance) are given to those patients who have a relatively high probability ($\geqslant 0.40$) of entering complete remission and an investigational treatment is given to those patients who have low predicted probability (<0.40) of having a complete remission. Currently, AMSA-OAP is the treatment for patients with low predicted probabilities of remission. Complete remission rates, length of remission, and survival will be compared against those for comparable groups of patients in a large historical series. This series consists of all previously untreated adult patients (over 15 years of age) with acute leukemia who received primary treatment with AD-OAP at the same institution between 1973 and 1979.

This is an example of a large historical control study in which the groups of patients receiving AD-OAP treatment in the current and historical series will provide a justification for comparing AMSA-OAP patients with patients from the historical control series who received AD-OAP. That is, it is expected that the AD-OAP patients in the current study will have comparable response and survival experience to similar patients in the historical control series. Patients receiving AMSA-OAP treatment will be compared with a group of patients receiving AD-OAP in the historical control series; comparability will be determined utilizing a logistic regression model to calculate predicted probabilities of entering complete remission. The study is still in progress.

DISCUSSION

Though this paper gives a history of selected clinical trials over the last 33 years, giving a proper perspective to the contributions of randomized and nonrandomized trials may require some speculation. The period from 1955 to the present has been one in which many randomized trials have been conducted as part of the cooperative groups program sponsored by the National Cancer Institute. Hence, on general grounds one might expect randomized studies to be overrepresented in relation to the advances made. In some cases, one might speculate on the results of a clinical trial, assuming that it was nonrandomized rather than randomized or vice versa.

The nonrandomized clinical trial is the clear winner when the development of new therapies for induction of remissions is considered. The discovery of methotrexate, prednisone, 6-mercaptopurine, cytosine arabinoside, adriamycin, combination chemotherapy, and the importance of the protected environment have all been derived first from the nonrandomized studies. Concerning maintenance treatments, the study by Freireich et al. [8] was a randomized study in which 6-MP was established as an effective maintenance treatment, though it had already been shown to be an effective remission induction treatment. Further, there were the nonrandomized aspects of the design that also demonstrated 6-MP to be an effective maintenance treatment.

Randomized comparative studies are generally considered the standard way for ranking treatments, though such studies may not always be conclusive, e.g., the COAP vs. OAP vs. DOAP study. The credibility of any study is enhanced when confirmatory studies lead to the same conclusions.

Determining patient characteristics related to prognosis and developing regression models quantitating the relationship of these features to prognosis has occurred in both randomized and nonrandomized studies. The major important element seems to be to have a well-studied group of patients by the same investigators with complete information of high quality on prognostic features, treatment programs, and outcome.

A concept of independent action of drugs and the importance of combining them to achieve high remission rates can be attributed to both randomized and nonrandomized studies; clearly, however, the concept of independent drug action was best elucidated in a randomized study, since the design features made it easier to analyze for evidence of independent drug action.

The concept of treating patients in a disease-free state was established by a randomized study of 6-MP vs. placebo in the remission maintenance phase. This concept has been extended to other types of clinical trials, namely adjuvant studies conducted when patients are in a disease-free stage, e.g., breast cancer. Since the result that 6-MP could maintain remission longer than placebo was also found by using each patient as his or her own control, one might speculate that the concept of treating patients in complete remission could equally have been found in a well-conducted nonrandomized study.

Utilization of regression models in clinical studies makes it possible to compare treatment programs, adjusting in a precise and quantitative manner for the characteristics of the patients on each treatment. Basically, this development might be expected to be of more benefit to nonrandomized than randomized studies because the former relies on the model for analysis and interpretation, whereas the latter generally relies on the randomization to have balanced prognostic features among treatment groups (though this can be supplemented by use of the regression model). Clinical trials, both randomized and non-randomized, are not isolated phenomena. Many ideas for studies in acute leukemia were developed based upon studies in animals, phamacological concepts developed first through animal studies, and other clinical studies. Consequently, no clinical trial can be considered completely definitive; confirmatory studies are important for both randomized and nonrandomized studies.

ACKNOWLEDGMENT
I would like to thank Dr. E.J Freireich for several very helpful discussions concerning this paper and for the opportunity to collaborate in the conduct of studies for more than 20 years.

This paper was written while the author was on leave at the Département de Biophysique et de Biomathématiques, Université Pierre et Marie-Curie, Faculté de Médecine Pitié-Salpêtrière, Paris.

REFERENCES

1. Farber S, Diamond LK, Mercer R, et al: Temporary remissions in acute leukemia in children produced by folic acid antagonist aminopterin. New Engl J Med 238:787–793, 1948
2. Pearson OH, Eliel LP: Use of pituitary adrenocorticotropic hormone (ACTH) and cortisone in lymphomas and leukemias. J Am Med Assoc 144:1349, 1950
3. Farber S, Downing V, Schwachman H, Toch R, Appleton R, Heald F, King JP, Feriozi D: The action of ACTH and cortisone in acute leukemia. In: *Proceedings of Second Clinical ACTH Conference*, Mote JR, Ed. Philadelphia: The Blakiston Company, 1951, p 251
4. Burchenal JH, Murphy ML, Ellison RR, Karnofsky DA, Sykes MP, Tan CTC, Leone LS, Craver LF, Dargeon HD, Rhoads CP: Clinical evaluation of a new antimetabolite 6-mercaptopurine, in the treatment of leukemia and allied diseases. Blood 8:965, 1953
5. Freireich EJ, Frei E III, Holland JF, Pinkel D, Selawry O, Rothberg H, Haurani F, Taylor R, Gehan EA: Evaluation of a new chemotherapeutic agent in patients with "advanced refractory" acute leukemia: Studies of 6-azauracil. Blood 16:1268–1278, 1960
6. Freireich EJ, Gehan EA, Sulman D, Boggs DR, Frei E: The effect of chemotherapy on acute leukemia in the human. J Chronic Dis 14(6):593–608, 1961
7. Frei E, Freireich EJ, Gehan EA, et al: Studies of sequential and combination antimetabolite therapy in acute leukemia: 6-mercaptopurine and methotrexate. Blood 18(4):431–454, 1961
8. Freireich EJ, Gehan EA, et al: The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy. Blood 21(6):699–716, 1963
9. Freireich EJ, Frei E III: Recent advances in acute leukemia. In: *Progress in Hematology*, Moore CV, Brown EB, Eds. New York: Grune and Stratton, 1964. pp 189–202
10. Karon MR, Freireich EJ, Frei E III: A preliminary report of vincristine sulphate—a new active agent for the treatment of acute leukemia. Pediatrics 30:791–796, 1962
11. Freireich EJ, Bodey GP, Harris JE, et al: Therapy for acute granulocytic leukemia. Cancer Res 27:2573–2577, 1967
12. Bodey GP, Gehan EA, Freireich EJ, Frei E: Protected environment-prophylactic antibiotic program in the chemotherapy of acute leukemia. Am J Med Sci 262(3):138–151, 1971
13. Rodriguez V, Bodey GP, Freireich EJ, McCredie KG, Gutterman JU, Keating MJ, Smith TL, Gehan EA: Randomized trial of protected environment-prophylactic antibiotics in 145 adults with acute leukemia. Medicine 57(3):253–266, 1978
14. Coltman CA, Bodey GP, Hewlett JS, Haut A, Bickers J, Balcerzak SP, Costanzi JJ, Freireich EJ, McCredie KB, Groppe C, Smith TL, Gehan EA: Chemotherapy of acute leukemia. A comparison of vincristine, cytarabine, and prednisone alone and in combination with cyclophosphamide and daunorubicin. Arch Intern Med 138:1342–1348, 1978
15. Gehan EA, Smith TL, Buzdar AU, Keating MJ, Freireich EJ: Knowledge acquisition from historical data: Application to breast cancer patients. Bull Cancer (Paris) 67:4, 1980

Impact of Randomized Clinical Trials on Therapy of Primary Breast Cancer:

The NSABP Overview

Howard E. Rockette, Carol K. Redmond,
Bernard Fisher, and Participating NSABP Investigators

NSABP Headquarters, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT: The history of treatment for early breast cancer, particularly as it relates to the most recent calendar decade, exemplifies the contribution that can be made to resolving therapeutic controversies by well-designed prospective randomized clinical trials (RCTs). In this article we provide an overview of the breast cancer clinical trials that have been conducted by the National Surgical Adjuvant Project for Breast and Bowel Cancers (NSABP) since 1971. These trials have had a strong impact in changing the clinical management of breast cancer over the past decade. Emphasis will be given to those aspects of the studies that demonstrate the advantages and/or necessity for RCTs to evaluate alternative therapeutic approaches to the management of breast cancer.

KEY WORDS: *randomized clinical trial, breast cancer*

INTRODUCTION

The major aim of the National Surgical Adjuvant Project for Breast and Bowel Cancers (NSABP) clinical trials has been to answer important biological and therapeutic questions relative to primary human breast cancer. In this regard, these trials have been concerned with two general treatment questions:

1. To what extent does the management of local-regional disease affect end results?
2. Can long-term adjuvant systemic therapy reduce recurrence and mortality in patients with no evidence of distant metastases at the time of initial diagnosis?

OVERVIEW OF SURGICAL TRIALS

The Halstedian concept of tumor spread was virtually unquestioned by surgeons in this country for over seventy-five years [1]. Halsted's belief in the orderly progression of tumors provided the rationale for the "radical mastec-

---

Received June 14, 1982; accepted July 1, 1982.

Address reprint requests to: Dr. Howard E. Rockette, Associate Professor, University of Pittsburgh, 914 Scaife Hall, Pittsburgh, PA 15261, USA.

Controlled Clinical Trials 3:209–225 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030209-17$02.75

tomy" as the preferred operation for breast cancer. This concept of disease progression implies that operable breast cancer is a local–regional disease and that the extent and nuances of operation are the primary factors influencing patient outcome. In spite of progressively better surgical techniques based upon anatomical principles, patient survival curves remained disappointingly unchanged for many years. Information concerning tumor biology emerging during the last 20 years has led to an alternative hypothesis that operable breast cancer is a systemic disease and that variations in local–regional therapy are unlikely to substantially affect survival [2].

By 1970 it was apparent that there was a need for a prospective study to provide support for one or the other of these hypotheses in order to resolve the controversy. Accordingly, in 1971 NSABP initiated a large scale RCT (protocol B-04) among investigators at 34 institutions in the United States and Canada. Between August of 1971 and August of 1974, 1765 patients were entered into a study having as a goal evaluation of these alternative viewpoints. Women meeting rigid protocol requirements who were clinically axillary node negative were randomized so that they received one of three local-regional treatment regimens. They either had a conventional radical mastectomy (RM), a total (simple) mastectomy followed by local–regional radiation (TM + R), or a total mastectomy alone (TM) with subsequent removal of axillary nodes only if they subsequently become positive. Similarly, clinically positive node patients were treated by RM or TM + R. The current average time on study is 100 months, with a range of 81–120 months.

For Protocol B-04 the long-term findings have indicated little or no differences in survival regardless of the local–regional treatments employed. Table 1 summarizes the disease-free survival (DFS) and survival at five years for patients on this protocol. Although there is a clear lack of any statistically significant differences, this is not the major concern in interpreting the data.

Table 1 Protocol B-04 Disease-Free Survival (DFS) and Survival by Treatment Group[a]

Study group	Number of patients	Percent DFS[b] 5 years	Percent survival 5 years
Clinically Negative			
RM	362	71 (212)	75 (256)
TM + R	352	74 (215)	75 (265)
TM	365	67 (189)	74 (272)
Clinically Positive			
RM	291	54 (128)	62 (180)
TM + R	294	50 (114)	58 (171)

[a] As of June 30, 1981.

[b] Number of patients at risk appear in parentheses.

In this protocol the attention focuses on the chance that small differences in treatments may not be detected due to biased comparisons, small patient numbers, or inadequate follow-up. Since even small systematic biases would be a serious matter in this protocol, randomization was essential to eliminate the need for concern in this regard. The relatively good survival and heterogeneity in patient populations necessitates large sample sizes if modest differences in survival are to be detected, while the chronicity of early breast cancer requires long term follow-up if that data are to be credible.

According to the "Halstedian Hypothesis," failure to remove all tumor foci would result in the inevitable occurrence of new clinically perceptible tumors as well as the enhancement of distant disease. Information from a variety of sources, however, indicates that despite a high incidence of multifocal lesions in both breasts of women with cancer, rarely is there evidence of two or more overt cancers in the same breast, synchronous bilateral tumors are uncommon, and the incidence of an asynchronous primary tumor in the uninvolved breast fails to approach the incidence of occult lesions detected by random biopsy or at autopsy. A new protocol (B-06) was developed to explore the importance of multicentricity in primary breast cancer. This is the only RCT in the United States and Canada to evaluate the efficacy of segmental mastectomy in conjunction with axillary dissection as compared to total mastectomy with axillary dissection. One half of the patients randomized to segmental mastectomy also receive radiation therapy postoperatively. At present, 1022 women have been entered into the trial. Considerably larger numbers are needed and no firm end results are available.

OVERVIEW OF SYSTEMIC THERAPY PROTOCOLS

In keeping with the view that breast cancer becomes a systemic disease early in its clinical history, NSABP in 1972 initiated the first of a sequence of clinical trials employing long-term (2-year) administration of chemotherapy. While the trials have been oriented toward determining clinical benefit, they were designed to test in an orderly fashion hypotheses derived from experimental models. The logical starting point for evaluating the worth of adjuvant chemotherapy in breast cancer seemed to be to implement clinical trials testing single agents. After having determined their effectiveness, it would be appropriate to proceed with the evaluation of multiple agents in a stepwise fashion.

Over 4000 patients have been entered into five studies (Table 2). They fulfill specific criteria of eligibility that are essentially similar in all trials and have been reported in detail elsewhere. They were treated at 73 NSABP member institutions in the United States and Canada. Investigators from the Eastern Cooperative Oncology Group also collaborated in Protocols B-05 and B-07. The initial investigation, B-05, was a double-blind randomized trial to evaluate the effectiveness of the single agent 1-phenylalanine mustard (P). In that study one half of the patients received P and the other half were administered a placebo. Following demonstrations of some early benefit from the single agent [3] and in keeping with our original plan, a second study (B-07), which compared P with P + 5-fluorouracil (PF), was carried out. Next, the effects of

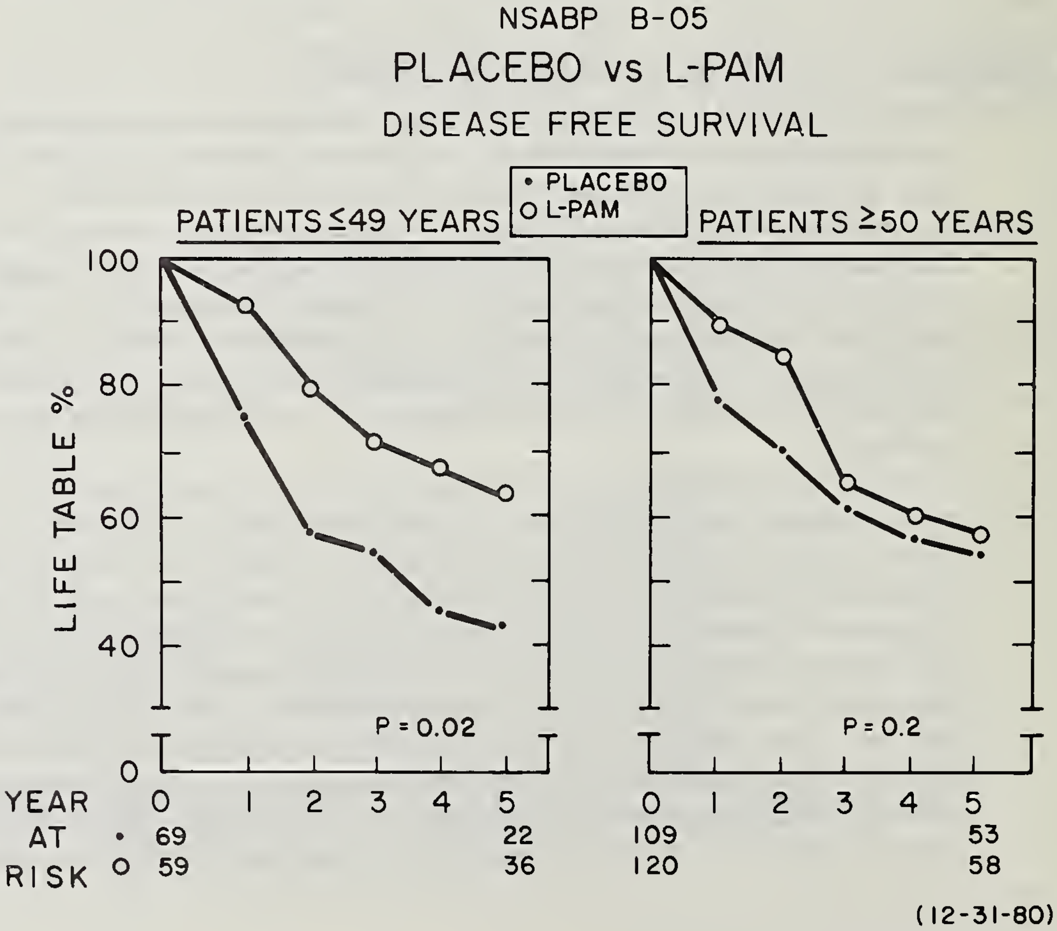
Table 2. NSABP Protocols Evaluating Systemic Therapy

Protocol		Randomization Begun	Terminated	Number of patients randomized (4003)	Average time on study (months)[a]
B-05	Placebo vs. P	9/27/72	2/5/75	370	88 (76–106)
B-07	P vs. PF	2/5/75	5/15/76	741	69 (61–77)
B-08	PF vs. PMF	4/12/76	4/29/77	737	56 (50–62)
B-09	PF vs. PFT	1/1/77	5/16/80	1891	31 (13–54)
B-10	PF vs. PF + CP	5/1/77	5/31/81	264	32 (5–50)

[a] As of June 30, 1981.

three drugs PF + methotrexate (PMF) were compared with those from PF (Protocol B-08). When early findings indicated that the two drug combination (PF) might be more beneficial in certain groups of patients than the single agent [4] and results were not yet available regarding the effectiveness of the three drugs, two additional studies were implemented. One (Protocol B-09)

Figure 1. Disease-free survival: placebo (•) vs. L-PAM (○).

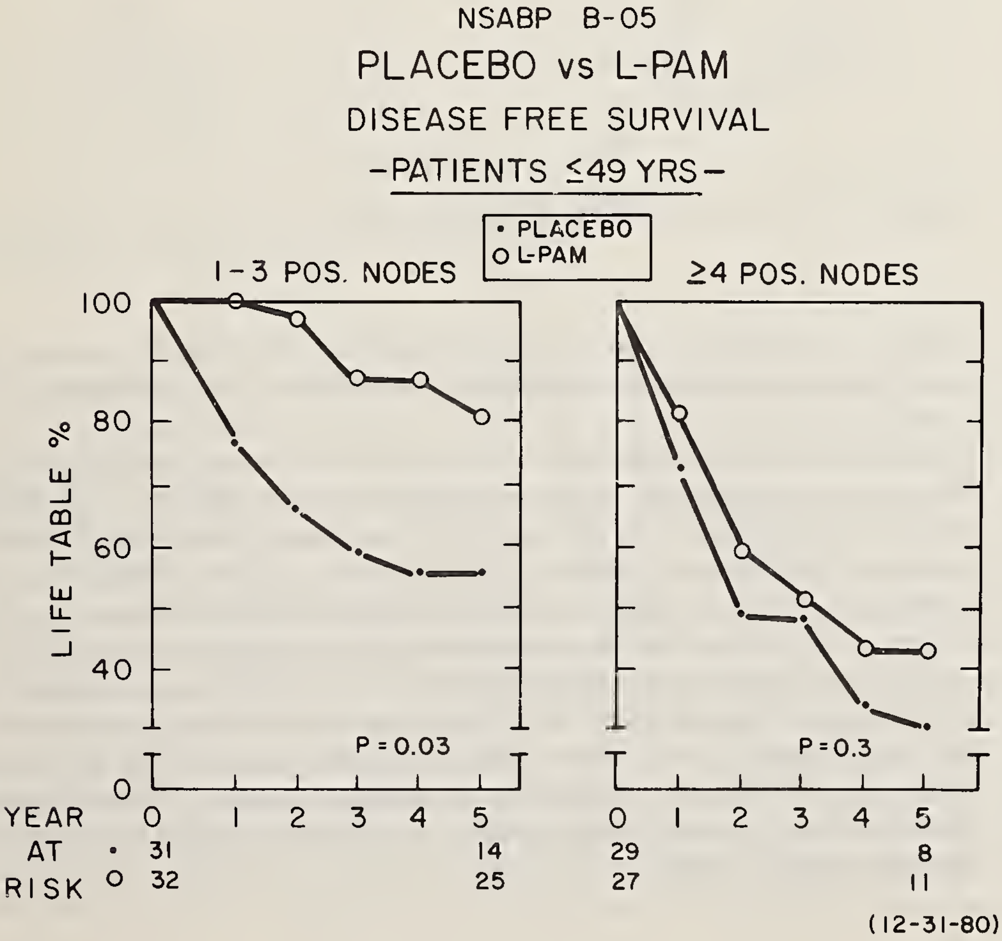


had as its objective the determination of whether the addition of tamoxifen (T) to PF would enhance disease free survival and survival. The other (Protocol B-10) was carried out to determine whether the addition to PF of the non-specific stimulating agent, *C. parvum*, was more beneficial than PF alone. (Since no data are yet published from the latter study, no further reference to it will be made.)

The early findings from Protocol B-05 indicated that adjuvant chemotherapy could alter the natural history of breast cancer, and that there were "subset" differences in response to chemotherapy; pre- and postmenopausal patients were affected differently. This was the first evidence supporting the use of adjuvant chemotherapy for at least a subset of patients aside from a prior NSABP study of single dose + Thio-TEPA conducted in the late 1950's [5].

In the first report of findings in 1975 it was observed that all patients who received P had a disease free survival significantly better than those receiving placebo ($p = 0.02$). Both pre- and postmenopausal women demonstrated a benefit. That difference in all patients has been found to persist 5 years later ($p = 0.03$). The early findings in premenopausal women have been maintained whereas the benefit noted in older women diminished after 2 years (Fig. 1). In patients ≤ 49 years the benefit has remained due primarily to the response to P on those with 1–3 positive nodes (Fig. 2). Although the survival pattern

Figure 2. Disease-free survival: placebo (•) vs. L-PAM (○).



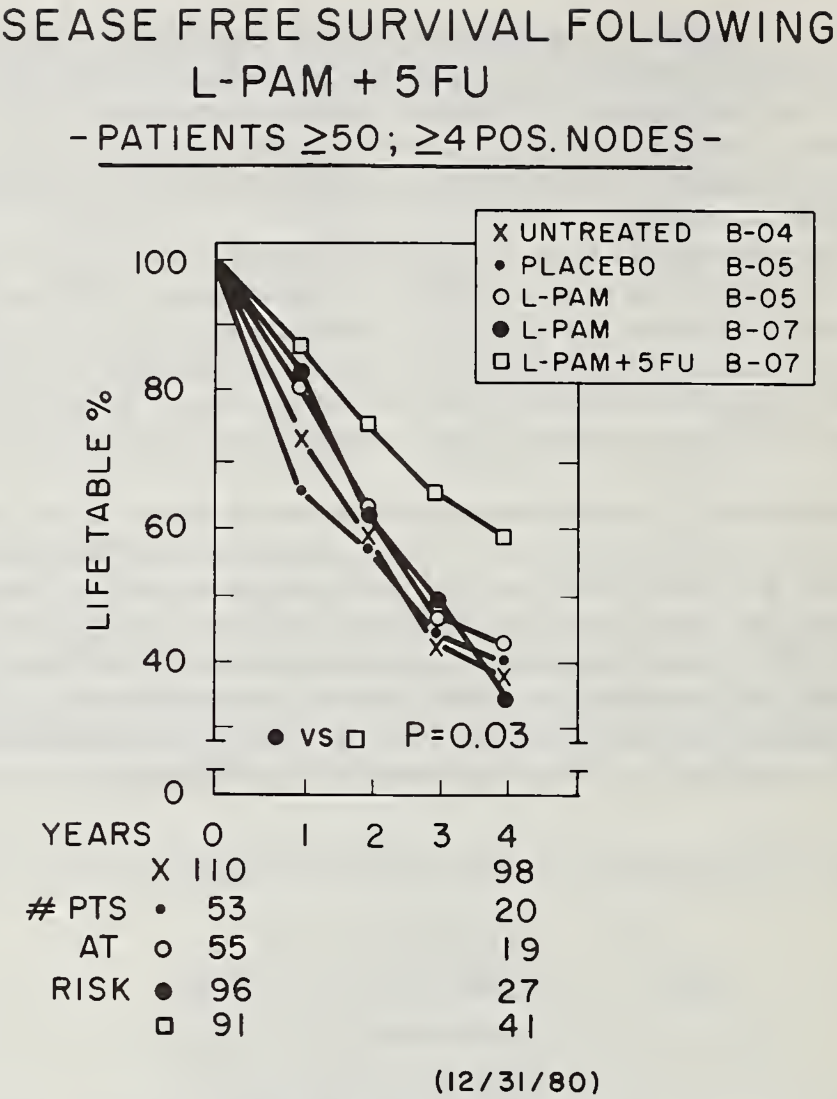


Figure 3. Disease-free survival following L-PAM + 5FU.

reflects the increased disease-free interval for patients on P, the small sample sizes result in a failure to show significant differences in survival at the present time.

Patients who received PF have been found to have an advantage in disease-free survival. Although there has been noted a consistent superiority for patients ≤ 49 years with the use of PF, the major difference in results were in those patients ≥ 50 years of age with four or more positive nodes. The latter group demonstrated a significant improvement in disease-free survival when compared to the patients treated with P alone ($p = 0.03$) or when compared to the same age-nodal group in other NSABP protocols that received no adjuvant chemotherapy or P alone (Fig. 3). The improvement in survival for this subgroup is consistent with the prolongment of the disease-free interval (Fig. 4). When patients are considered overall, according to age, or according to age and nodal status, findings to date have not indicated that PMF is superior to PF.

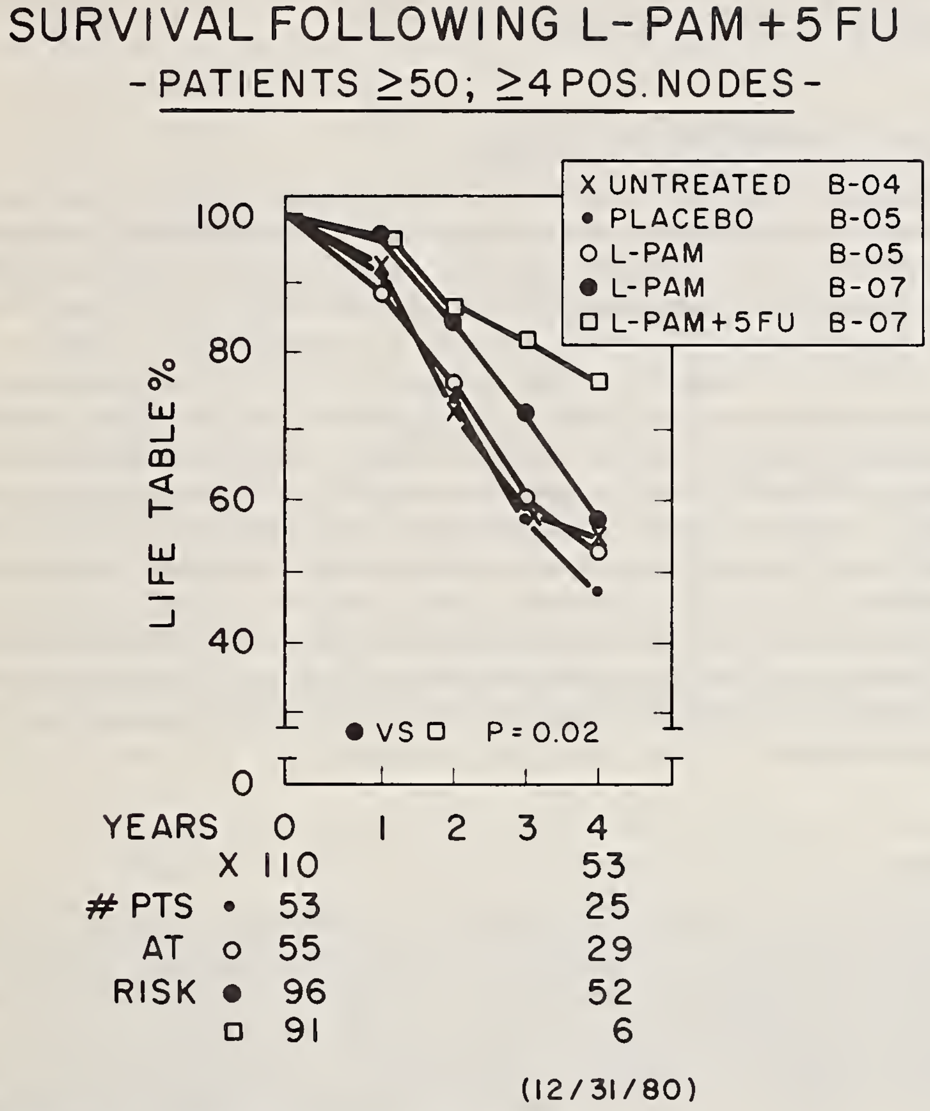


Figure 4. Survival following L-PAM + 5FU.

The antitumor effect of tamoxifen (T), a synthetic nonsteroid compound having antiestrogenic activity, in patients with metastatic breast cancer suggested its possible value in women with primary disease [6–8]. The primary objective of B-09 was to determine whether the addition of T to PF would enhance disease-free survival of PF treated patients. Preliminary results out to 2 years have recently been published for this protocol [9]. The overall results show a statistically significant difference in disease-free survival in favor of PFT. Analyses relative to the estrogen receptor (ER) and progesterone receptor (PR) proteins in the primary tumor indicate that there is a strong interaction of response to treatment with age or age related factors. In patients under 50 years of age having either ER or PR receptor poor tumors (<10 fmol) who received PFT, both disease-free survival and survival were significantly lower than those treated with PF. In patients 50–59 years of age with low PR, the response to treatment closely resembled the response in patients ≤ 49 years, while for patients with high PR the response was similar to patients 60–70 years of age. In patients 60–70 years of age a benefit from PFT has been noted regardless of ER and PR status.

NEWLY IMPLEMENTED NSABP STUDIES

In keeping with NSABP policy of designing new protocols based upon information currently available from our previous studies and progressing in sequential fashion, two new trials for breast cancer patients with positive nodes have been implemented. Patients receive PF or PF plus adriamycin (PAF) on Protocol B-11 and on Protocol B-12 receive PFT or PAFT.

For the first time the NSABP has implemented chemotherapeutic trials utilizing negative node patients. To the present, such evaluations have not been undertaken because of the lack of a discriminant that could separate out those patients who have the greatest probability of demonstrating a treatment failure. Moreover, there has been concern regarding the use of an alkylating agent in patients who overall have a favorable prognosis. Recent evidence indicating that the ER of tumors in node negative patients serves as such a discriminant has led us to proceed with two trials. Those whose tumors are ER negative will either receive placebo or will receive methotrexate followed by F (M → F). Estrogen receptor positive patients will receive either placebo or tamoxifen.

COMMENTS

While the controversies surrounding the appropriate treatment for primary breast cancer are not all resolved, there has been continuing progress made throughout the past 10 years. Final evaluation of many of these trials will require additional observation time. However, it is important to note that the natural history of the disease has been altered by the chemotherapy employed as an adjuvant in the NSABP trials. Differences in survival have persisted for out to 5 years or more at this time. Even if in future follow-up these gains are diminished or lost, improvement in survival up to 5 years or longer seems sufficient to conclude that chemotherapy is worthwhile.

It seems unlikely that such substantial progress could have been made outside of large-scale RCTs. The advances that have been made have resulted in small to moderate increments in survival or disease-free survival across the adjuvant therapy trials. The heterogeneity of the patient population relative to other factors related to disease-free survival make it difficult to determine the treatment effect. In addition to the need for large numbers of patients, such effects can only achieve credibility within the framework of randomized trials to rule out any systematic biases that could otherwise distort these small differences. Since new prognostic factors continue to be identified their effect cannot be evaluated by post stratification. A good example is the relationship of hormone receptor proteins to the end results on hormonal therapy. Since information on such receptor proteins was not available for earlier clinical trials in breast cancer, the use of historical or nonrandomized controls might have resulted in biased comparisons.

The need for randomization in the investigation of early stage breast cancer may be more crucial than for cancer of other sites. Competing treatment regimens have been widely publicized and discussed by the media and potential patients form opinions relative to treatment. In a nonrandomized setting patient selection of treatment may result in incorrect conclusions if the treatment selection is related to patient variables which affect end results. Furthermore, the extensive use of screening for identification of breast cancer has resulted in changes in the distribution of the prognostic factors of patients at the time of diagnosis. This relationship of patient characteristics to calendar time makes the selection of a historical control particularly difficult.

Relative to clinical trial methodology, our experience has indicated to us a number of aspects that we believe are essential for such endeavors to be successful in establishing effective cancer therapies:

(1) There is a need for larger sample sizes than most clinical trials have employed for adjuvant Phase III clinical studies. The use of sample size considerations based upon number of failures expected in a homogeneous population are inadequate in a population as heterogeneous as breast cancer patients. Further, there is considerable evidence in breast cancer that not all patient groups will respond similarly. Important biological information can result if adequate numbers are provided in the study to sort out these heterogeneous treatment effects. While statistical modeling may do much to assist in understanding the interrelationships among prognostic variables and end results, there is no substitute for large numbers of patients to provide conclusive results.

(2) Because of the relatively good prognosis of most breast cancer patients, long-term follow-up is necessary. The problems encountered in obtaining accurate follow-up become increasingly difficult for patients as their time from entry into the study increases. It is not sufficient to follow patients merely to first recurrence of disease since survival provides important additional information regarding the influence of adjuvant chemotherapy on success of other treatments employed at the time of recurrence. It is possible that with certain therapies the gains achieved in disease-free survival may not subsequently be reflected in survival.

(3) The need for large patient numbers necessitates involving many institutions in each trial. Complex therapeutic schemes with large numbers of treatments employed simultaneously cannot usually be carried out in diverse settings. Since quality control is more difficult, it is necessary to develop straightforward study aims with reasonable data collection, if the study is to be of appropriate quality. Because of the constant turnover that occurs at institutions, one must refamiliarize individuals with the protocol by careful checking of eligibility at the time of randomization and frequent reminders of information due.

There are some advantages to trials involving many institutions that compensate to some extent for the additional difficulties that may be entailed with data collection and quality control. A general consistency across the institutions in their agreement with the end results generally provides credibility to the findings of the study. The population to which the results may be generalized is broadened and one has a good idea by the end of the study of the extent to which the particular regimen can be successfully used across diverse clinical settings.

(4) Finally, there is a need to plan not just single clinical trials but rather a general approach to a disease predicated upon some understanding of its natural history. We feel that the NSABP protocols have been important not merely from the standpoint of comparing treatment A to treatment B, but in providing important biological information about the disease. The use of a systematic approach to planning clinical trials building upon all the experience from previous trials is important if progress is to be made in other solid tumors.

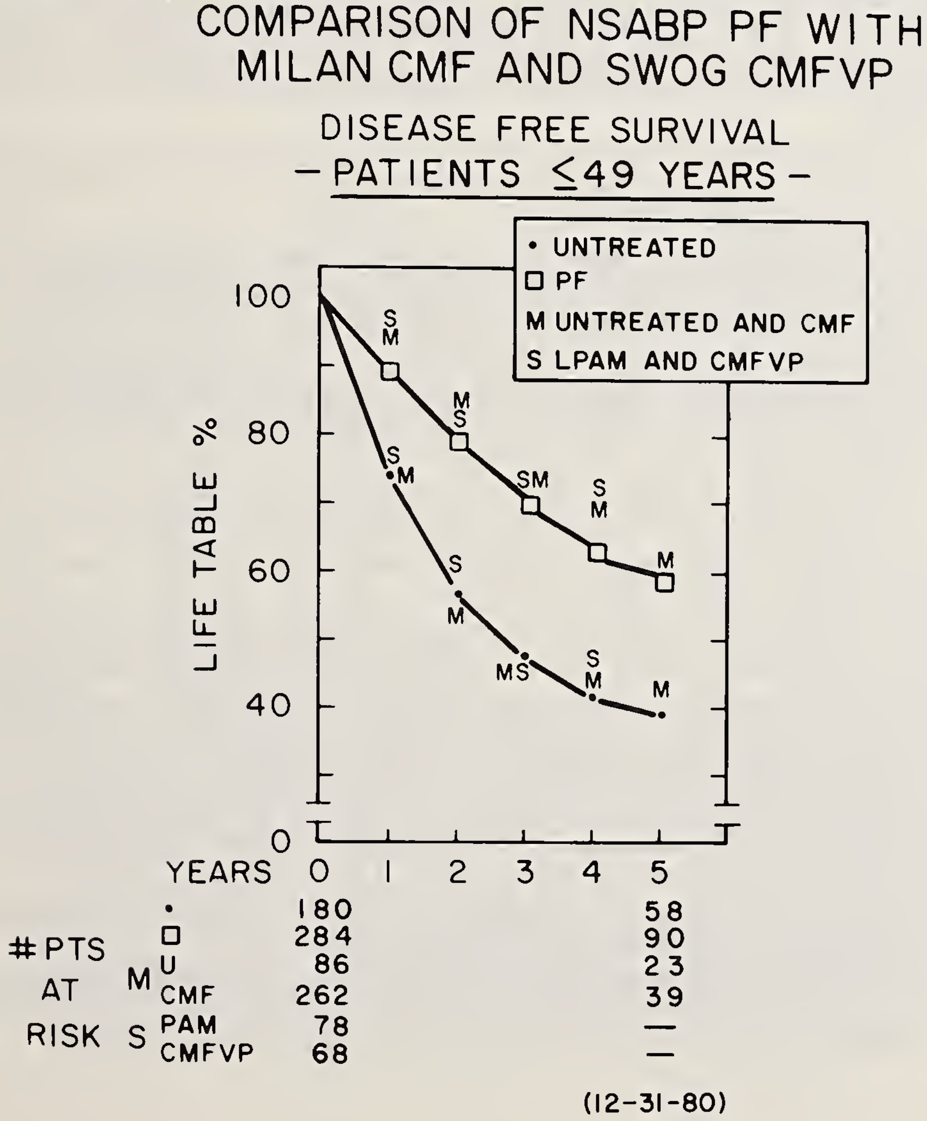
One aspect of NSABP protocols that differs from many others that have been conducted is the sequential systematic approach to subsequent protocols. The first chemotherapy protocol was one drug versus placebo, followed by a second protocol of two drugs versus the one drug appearing as the treatment arm of the previous protocol, three drugs versus two drugs where the two-drug protocol was the treatment arm of the preceding protocol. This sequential buildup seems desirable when using drugs where toxicity considerations are important. From our findings it appears that the potential always exists to build on earlier successes. Thus an increase in survival obtained with one drug may be built upon by additional therapy. However, due to the toxicity of the various drugs, if one starts with a multiple chemotherapy regimen with high toxicity, when success occurs, the ability for further improvement by additional adjuvant therapy is limited. The alternative is to then try to regress and eliminate chemotherapy to determine what is the minimal subset of drugs required to achieve the same result. Another advantage of the sequential buildup is that the treatment arm of each protocol appears as a control arm in a subsequent protocol and this allows more flexibility relative to comparisons among different protocols. Although such comparisons are not the same as a randomized control group, we feel that these comparisons across protocols with similar eligibility are much better than a typical historical control.

Although progress has been made in the treatment of breast cancer, there remain a number of important questions that require continuation of existing or initiation of new randomized clinical trials. While Protocol B-04 and the recent study of quadrantectomy reported by the Milan group indicates that lesser surgical procedures may be considered for certain groups of primary breast cancer patients, the question regarding whether "lumpectomy" is equivalent to breast removal remains an important issue addressed in our Protocol B-06. Such a study, in which large numbers are required in order to demonstrate that patients are not harmed through receiving the lesser surgical or surgical plus radiation treatment, cannot be resolved except within the framework of the RCT.

Adjuvant chemotherapy has been demonstrated to improve disease-free survival and survival, but the use of more than two chemotherapeutic agents has not been shown conclusively to achieve additional benefits. Resolving the issue of which of the current chemotherapeutic regimens available is most beneficial can be considered important since the use of additional agents results in increased toxicity to the patient. The PF regimen that has been shown to be most effective in improving survival in NSABP protocols to date has been administered to large numbers of patients with a correspondingly low incidence of severe or life-threatening acute toxicity. On the other hand, out to over 4 years of observation we have not been able to demonstrate any improvement of disease-free survival or survival for any major subgroup of patients receiving the PMF regimen. To the limited extent that the literature permits comparisons with other recent trials, it is not apparent that the three-drug CMF (Milan) [10,11] or five-drug CMFVP (SWOG) [12] regimens have demonstrated any additional gains in disease-free survival over PF (Fig. 5). Although survival data are not completely available from these studies, what is published suggests that the end results achieved by PF are at least as good as those achieved by the greater number of drugs.

The systematic buildup in NSABP protocols utilizing as additional drugs those that have been shown to have some effect in the advanced disease trials will, we feel, result in eventually culling out the most beneficial of the drug regimens with appropriate regard for any toxic effects. It remains to be demonstrated whether the increased disease-free survival that has been noted for postmenopausal patients receiving adjuvant chemotherapy plus the anties-

Figure 5. Comparison of NSABP PF with Milan CMF and SWOG CMFVP.



trogen tamoxifen will result in a corresponding benefit in their survival experience.

In summary, it appears that substantial progress has been made in the treatment of early-stage primary breast cancer utilizing RCTs and current trials should add to this progress in the near future. Problems that exist relative to randomized clinical trials can be related more to the way such trials have been designed and executed rather than to criticisms of the methodology itself, which when properly carried out has been demonstrated to provide important new information. Further, the methodology and sophistication in its application are continually improving over time. When ethical issues do not preclude its use, the appropriate focus should be upon how the principles may be best utilized rather than upon what the alternative approaches to the RCT might be.

REFERENCES

1. Halsted WS: The results of radical operations for the cure of carcinoma of the breast. Ann Surg 46:1, 1907.
2. Fisher B, Redmond C, Fisher ER, Participating NSABP Investigators: The contribution of recent NSABP clinical trials of primary breast cancer therapy to an understanding of tumor biology—an overview of findings. Cancer 46:1009–1025, 1980.
3. Fisher B, Carbone P, Economou SG, et al: L-phenylalanine mustard (L-PAM) in the management of primary breast cancer: A report of early findings. New Engl J Med 292:117–122, 1975.
4. Fisher B, Glass A, Redmond C, et al: L-phenylalanine mustard (L-PAM) in the management of primary breast cancer: An update of earlier findings and a comparison with those utilizing L-PAM plus 5-fluorouracil (5-FU). Cancer 39:2883–3903, 1977.
5. Fisher B, Ravdin RG, Ausman RK, et al: Surgical adjuvant chemotherapy in cancer of the breast: Results of a decade of cooperative investigation. Ann Surg 168:337–356, 1968.
6. Cole MP, Jones CTA, Todd IDH: A new anti-oestrogenic agent in late breast cancer: An early clinical appraisal of ICI 46474. Brit J Cancer 25:270–275, 1971.
7. Ward HWC: Anti-oestrogen therapy for breast cancer: A trial of tamoxifen at two dose levels. Brit Med J 1:13–14, 1973.
8. Lerner HJ, Band PR, Israel L, Leung BS: Phase II study of tamoxifen: Report of 74 patients with stage IV breast cancer. Cancer Treat Rep 60:1431–1435, 1976.
9. Fisher B, Redmond C, Brown A, et al: The treatment of primary breast cancer with chemotherapy and tamoxifen: Correlation of quantitative estrogen receptor levels with response to therapy. New Engl J Med 305:1–6, 1981.
10. Bonadonna G, Valagussa P: Adjuvant therapy of primary breast cancer. In: *Principles of Cancer Treatment*, Carter SK, Glatstein E, Livingston RB, Eds. (to be published).
11. Bonadonna G, Valagussa P: Dose response effect of adjuvant chemotherapy in breast cancer. New Engl J Med 304:10–15, 1981.
12. Glucksberg H, Rivkin SE, Rasmussen S, et al: Combination chemotherapy (CMFVP) versus L-phenylalanine Mustard (L-PAM) for operable breast cancer with positive axillary nodes: A southwest oncology group study. Cancer (in press).

APPENDIX 1: LIST OF NSABP PROTOCOL NO. 4 PARTICIPANTS

NSABP institution	Responsible surgeon	Responsible radiation oncologist	Responsible pathologist
Albert Einstein College of Medicine	Herbert Volk	N.A. Ghossein	John Molnar
Boston City Hospital (Tufts)	C. William Kaiser	Merrill Feldman	Leonard Berman
Creighton University	Claude H. Organ	J.R. Zastera	Wade Bardawil
Downstate (SUNY)	Bernard Gardner	Joseph Bohorquez	Yale Rosen
Ellis Fischel State Cancer Hospital	William Donegan	James Thomson	Carlos M. Perez-Mesa
Fitzsimons General Hospital	Richard M. Hirata	Terry Powell	William Starke
French and Polyclinic Medical School	James McManus	George Schwarz	William Finkelstein
Geisinger Medical Center	C.W. Konvolinka	David Beiler	C. James Favino
Harbor General Hospital (UCLA)	John R. Benfield	Richard Small	Frank M. Hirose
Harrison S. Martland Hospital	Benjamin F. Rush, Jr.	John Mallams	Michael Lyons
Hennepin County General Hospital	Claude R. Hitchcock	Manoucher Azad	John I. Coe
Jewish General Hospital	Richard G. Margolese	J.J. Hazel	Claude LaChance
Louisiana State University	Isidore Cohn, Jr.	Joseph V. Schlosser	Ronald A. Welsh
Medical College of Pennsylvania	Donald Cooper	Janet Parker	Gerald Justh
Medical College of Virginia	Walter Lawrence, Jr.	E. Richard King	Saul Kay
Metropolitan Hospital	John F. Weiksnar	William Henkin	Eugene Schwartz
Michael Reese Hospital	Richard H. Evans	Lionel Cohen	Miriam L. Christ
Montefiore Hospital	Richard G. Rosen	Flora Mincer	Norwin Becker
Mount Sinai (New York)	Gerson J. Lesnick	John Boland	Mamoru Kaneko
St. Vincent's Hospital (New York)	Thomas Nealon, Jr.	George Schwarz	William E. Delaney, III
Temple University	Willis Maier	Marie LoPonte	Paul Putong
University of Arkansas	Kent Westbrook	Eleanor Deed	H.K. Leathers
University of California (San Diego)	Marshall J. Orloff	Carl Von Essen	Sidney Saltzstein
University of Illinois	Tapas K. Das Gupta	Edwin J. Liebner	Jose Manaligod
University of Iowa	Richard L. Lawton	Howard B. Latourette	Frederick W. Stamler
U.S. Naval Hospital (San Diego)	T. James Guzik	Quintus Crews	Frances Wachter
University of Pennsylvania	Francis E. Rosato	Lawrence Davis	Horatio T. Enterline
University of Pittsburgh	Bernard Fisher	John Parsons	Robert Totten
University of Rochester	W. Bradford Patterson	Eileen Paterson	Robert Cooper, Jr.
University of Texas (San Antonio)	Anatolio B. Cruz, Jr.	Peter Zanca	George Bannayan
University of Texas (Galveston)	Edward B. Rowe	Marvin H. Olson	Richard Marshall
University of Vermont	Roger S. Foster	G. Stephen Brown	David Duffell
Washington University	Harvey Butcher	Carlos A. Perez	Walter Bauer
Wayne State University	Alexander J. Walt	Alan Scheer	Barbara Rosenberg

APPENDIX 2: LIST OF NSABP PARTICIPANTS IN PROTOCOLS B-05, B-07 AND B-08

NSABP institution	Current surgeon or medical oncologist	Current pathologist
Akron City Hospital	Marvin Sakol	L.B. Reyes
Albert Einstein (NY)	Herbert Volk	Boyce Bennett
Boston City Hospital	Arthur Glasgow	Leonard Berman
Boston University	Merrill Feldman	Philip Gordon
Daniel Freeman Hospital (CA)	Elliott Hinkes	Raouf Yuja
Dekalb General Hospital (GA)	S. Angier Wills	Frank Matthews
Denver General Hospital	George Moore	Joseph Preston
Downstate (S.U.N.Y.)	Bernard Gardner	Anthony Nicastri
Ellis Fischel (MO)	William Johnston	Carlos Perez-Mesa
French Polyclinic (NY)	James McManus	Nijole Brazenas
Georgia Baptist Hospital	A. Hamblin Letton	Lester Forbes
Greater Bakersfield (CA)	James E. Donovan	Robert Wybel
Group Health Med. Ctr. (WA)	Robert Bourdeau	Charles Marshall
Hackensack Hospital (NJ)	Bernard Koven	Stuart Anderson
Harbor General Hospital (CA)	David State Yosef Pilch	Frank Hirose
Hennepin Co. Gen. Hosp. (MN)	Claude Hitchcock	John Coe
Hershey Medical Center (PA)	William DeMuth	Samuel Ward
Highland Hospital (NY)	Edwin Savlov	David Platt
Hotel-Dieu (Montreal)	Jacques Cantin	Yvan Boivin
Hotel-Dieu (Quebec City)	Louis Dionne	Real Lagace
Huron Road Hospital	M.D. Ram	Edward Siegler
Jewish General Hospital	Richard Margolese	Claude LaChance
Kaiser Permanente (Fontana, CA)	Vinod Patwardhan	Roy Hoke
Kaiser Permanente (Harbor City)	Eugene Pollack	John Rowe
Kaiser Permanente (Portland, OR)	Andrew Glass	Tsoi Tsuen Chan
Kaiser Permanente (San Diego)	Thomas Campbell	Marvin Nicola
Letterman General Hospital (CA)	Glenn Justice	Philip Stansifer
Louisiana State University	Isidore Cohn	Ronald Welsh
Lynn Hospital (MA)	Bernard Willett	Emei Shen
Medical College of Virginia	Walter Lawrence	Saul Kay
Medical College of Wisconsin	William Donegan	Lawrence Clowry
Memorial Ca. Res. Fdtn. (CA)	David Plotkin	Victor Rosen
Mercy Hospital (IA)	Albert Clemens	Joseph Song
Metropolitan Hospital	John Weiksnar	Eugene Schwartz
Michael Reese Hospital	Richard Desser	Wendy Recant
Michigan State University (East Lansing, MI)	Leif Suhrland	Ralph Edminster
Montefiore Hospital (NY)	Richard Rosen	Norwin Becker
Montreal General Hospital	John MacFarlane	W.P. Duguid
Mount Sinai (Cleveland)	Sidney Sachs	E. Siegler
Newark Beth Israel (NJ)	Frederick Cohen	Alkiviadis Campbell
Oakland Naval Hospital (CA)	Captain R.M. Deaner	Philip Vogt
Pennsylvania Hospital	Harvey Lerner	Santo Longo
Roger Williams Gen. Hospital (RI)	Jack Savran Frank Cummings	Israel Diamond
Royal Victoria Hosp. (Montreal)	Henry Shibata	Giles Tremblay
Rush-Presby. St. Luke's (IL)	Steven Economou	Edmund Pellettiere
Saginaw Cooperative Hosp. (MI)	Nikolay Dimitrov	John Young
South Nassau Community Hosp. (NY)	Nicholas LiCalzi	Ahmed Khapra

APPENDIX 3: INSTITUTIONS ACTIVELY CONTRIBUTING TO NSABP PROTOCOLS

NSABP institutions[a]	Current principal investigator
Albany Regional Cancer Center, Albany, NY	Thomas Cunningham
Albert Einstein College of Medicine, Bronx, NY	Herbert Volk
Albert Einstein Medical Center, Philadelphia, PA	Stanley Levick
Allentown Hospital, Allentown, PA	David Prager
Berkshire Medical Center, Pittsfield, MA	Jesse Spector/Harvey Zimbler
Boston University, Boston, MA	Peter J. Deckers/Merrill Feldman
Cross Cancer Institute, Edmonton, Alberta	Sandy Paterson
Daniel Freeman Hospital, Inglewood, CA	Elliott Hinkes
Dekalb General Hospital, Decatur, GA	S. Angier Wills
Denver General Hospital, Denver, CO	George Moore
Downstate Medical Center, Brooklyn, NY	Bernard Gardner
Ellis Fischel State Cancer Hospital, Columbia, MO	William Kraybill
Geisinger Medical Center, Danville, PA	Philip Breen
Group Health Medical Center, Seattle, WA	Robert Bourdeau
Harbor General Hospital, Torrance, CA	David State/M. Michael Shabot
Hennepin County Medical Center, Minneapolis, MN	Claude Hitchcock
Highland Hospital/University of Rochester, Rochester, NY	Sidney Sobel
Hopital Avicenne, Paris, France	Lucien Israel
Hotel-Dieu, Quebec City, Quebec	Louis Dionne
Jewish General Hospital, Montreal, Quebec	Richard Margolese
Kaiser Permanente, Harbor City, CA	Eugene Pollack
Kaiser Permanente, Portland, OR	Andrew Glass
Kaiser Permanente, San Diego, CA	Thomas Campbell
Letterman Army Medical Center, San Francisco, CA	David Gandara
Louisiana State University, New Orleans, LA	Isidore Cohn
Louisiana State University, Shreveport, LA	Leonard Goldman
Manitoba Cancer Foundation, Winnipeg, Manitoba	David Bowman
Marin General Hospital, San Rafael, CA	Peter Eisenberg
McMaster University, Hamilton, Ontario	S.E. O'Brien
Medical College of Virginia, Richmond, VA	Walter Lawrence
Medical College of Wisconsin, Milwaukee, WI	William Donegan
Memorial Cancer Research Foundation, Culver City, CA	David Plotkin
Michael Reese Hospital, Chicago, IL	Richard Desser
Michigan State University, East Lansing, MI	Nikolay Dimitrov/Leif Suhrland
Montefiore Hospital & Medical Center, Bronx, NY	Richard G. Rosen
Montreal General Hospital, Montreal, Quebec	John MacFarlane
Mount Sinai Hospital, Cleveland, OH	Sidney Sachs
Newark Beth Israel Hospital, Newark, NJ	Frederick Cohen
Ottawa Civic Hospital/Ontario Cancer Fndt, Ottawa	James Devitt/Rebecca McDermot
Pennsylvania Hospital, Philadelphia, PA	Harvey Lerner
Pondville Hospital, Walpole, MA	Leo Stolbach
Royal Victoria Hospital, Montreal, Quebec	Henry Shibata
Rush Presbyterian-St. Luke's Medical Center, Chicago, IL	Steven Economou/Janet Wolter
St. Luc Hospital, Montreal, Quebec	Roger Poisson/S. Legault-Poisson
St. Luke's Hospital, Kansas City, MO	Paul Koontz
St. Michael's Hospital, Toronto, Canada	Leo Mahoney
St. Vincent's Hospital, New York, NY	Thomas Nealon
South Nassau Communities Hospital, Oceanside, NY	Nicholas LiCalzi
Southern Alberta Cancer Centre, Calgary, Alberta	L. Martin Jerry
Tulane University, New Orleans, LA	Carl Sutherland
U.S. Naval Hospital, Oakland, CA	James Congdon
U.S. Naval Hospital, San Diego, CA	Jim Guzik

APPENDIX 3 *(continued)*

NSABP institutions[a]	Current principal investigator
University of California, San Diego, CA	Yosef Pilch
University of Florida (JHEP), Jacksonville, FL	Neil Abramson
University of Hawaii, Manoa, HI	Noboru Oishi
University of Iowa, Iowa City, IA	Peter Jochimsen
University of Louisville, Louisville, KY	Joseph C. Allegra
University of Maryland, Baltimore, MD	E. George Elias
University of North Carolina, Chapel Hill, NC	Robert Capizzi
University of Pittsburgh, Pittsburgh, PA	Bernard Fisher
University of Texas, San Antonio, TX	A. B. Cruz/J. Bradley Aust
University of Vermont, Burlington, VT	Roger Foster
Valley Hospital, Ridgewood, NJ	Hugh Auchincloss
Washington Regional Medical Center, Fayetteville, AR	Arthur Hoge
Washington University, St. Louis, MO	Marc Wallack
West Suburban Hospital, Oak Park, IL	Everett Nicholas
White Memorial Medical Center, Los Angeles, CA	Matthew Tan
Wilmington Medical Center, Wilmington, DE	Robert Frelick

[a] Sixty-eight institutions.

APPENDIX 4: LIST OF NSABP PARTICIPANTS IN PROTOCOLS B-05, B-07 AND B-08

NSABP institution	Current surgeon or medical oncologist	Current pathologist
St. Luc Hospital (Montreal)	Roger Poisson	Claude Larose
St. Luke's Hospital (MO)	Paul Koontz	John Rippey
St. Vincent's Hospital (IN)	John Cavins	Lee Foster
St. Vincent's Hospital (NY)	Thomas Nealon	Ilona Toth
Tulane University	Edward Krementz Carl Sutherland	H. Nina Dhurander
University of Arkansas	Kent Westbrook	Glenn Baker
University of California, San Diego	Yosef Pilch	Sidney Saltzstein
University of Illinois	Tapas Das Gupta	Jose Manaligod
University of Iowa	Richard Lawton Peter Jochimsen	Frederic Stamler
University of Louisville	Carl Knutson Condict Moore	William Christopherson
University of Maryland	E. George Elias	Cyril Toker
University of Mississippi	George Smith	Carl Evers
University of Pittsburgh	Bernard Fisher	Susan Rogers Edwin R. Fisher
University of Rochester	W. Bradford Patterson	David Platt
University of Texas, San Antonio	J. Bradley Aust Anatolio B. Cruz	George Bannayan
University of Vermont	Roger Foster	Roy Korson
University of Wisconsin	Douglas Tormey	Kennedy Gilchrist
USC-LA County Medical Center	Lewis Guiss	Peter Schwinn
U.S. Naval Hospital (CA)	T. James Guzik	Adelbert Cramer
Valley Hospital (NJ)	Hugh Auchincloss	William O. Green
Washington University	Harvey Butcher Gordon Philpott	Walter Bauer

APPENDIX 4 *(continued)*

NSABP institution	Current surgeon or medical oncologist	Current pathologist
Wayne State University	Alexander Walt	Barbara Rosenberg
West Suburban Hospital (IL)	Everett Nicholas	Frederick Volini
White Memorial Med. Center (CA)	Samuel Fritz	George Kypridakis
Wilmington Medical Center	Robert Frelick	Patrick Ashley
ECOG		
Allentown Hospital (PA)	David Prager	Paul Panas
Baptist Hospital (FL)	Edward Knight	Albert Drlicka
Berkshire Medical Center (MA)	Jesse Spector	Alex Parkinson
Cleveland Metropolitan Hospital	Edward Mansour	John Reid
Chicago Medical School	Ediz Ezdinli	Osualdo L. Rubinstein
Georgetown University (D.C.)	Bruce Schnider	Dae Sok Koh
Jefferson Hospital	Arthur Weiss	Arthur Patchefsky
Jefferson Medical College	Harvey Brodovsky	Arthur Patchefsky
Lahey Clinic Foundation	Richard Oberfield	Melvin Legg
Ontario Cancer Foundation	David Klaassen	E. Lieta
Paris	Lucien Israel	Coston Brard
Tufts-Pondville Hospital (MA)	Leo Stolbach	Carol Amich
University of Alberta	Pierre Band	Gordon Bain
University of Manitoba	Martin Levitt David Bowman	M.P. Hamonic Donald Penner
University of Pennsylvania	John Glick	Horatio Enterline
COG		
Duke University (NC)	William Shingelton	Bernard Fetter
Ohio State University	John Minton	William Holaday

Assessment of the Role of Randomized Clinical Trials in Establishing Treatment Policies

H. S. Garnier, R. Flamant, and C. Fohanno

Department of Medical Statistics, Gustave-Roussy Institute, Villejuif, France

Randomized clinical trials (RCTs) have now been performed for more than 30 years and are becoming increasingly numerous. Cooperative groups consisting of physicians, statisticians, and computer specialists have been established in several countries to undertake such trials. They necessitate a strong administrative structure and often the creation of a data center, entailing high costs.

Thus, it appears essential to assess the role of RCT in the improvement of therapeutic results and to determine whether the benefits outweigh the expenses incurred. The success of such a task, however, is not certain.

Oncology is a field where many such trials have been conducted, and the experience that can be drawn from this discipline should be profitable to other medical fields as well. There are at least two ways of evaluating the impact of RCTs on therapeutic progress:

1. considering the randomized trials already completed and measuring the role of their results, whether positive or negative, in the change of therapeutic policies in subsequent years; and
2. considering the treatments held as the most effective by the medical community and used as standards, and evaluating the respective roles of randomized and non randomized clinical trials in the establishment of those treatments.

SURVEY OF THE RANDOMIZED TRIALS IN CANCER

In ten years (1968–78), the Information Office of the International Union Against Cancer (UICC) registered approximately one thousand trials performed in over 30 different countries [1]. Some of these trials have already been completed and published. An annual questionnaire made it possible to follow the progress of the trials, to trace any trial that was abandoned, or not

---

Received June 14, 1982; revised July 1, 1982.

Address reprint requests to: Dr. C. Fohanno, Department of Medical Statistics, Gustave-Roussy Institute, 39 à 53, rue Camille Desmoulins, 94805 Villejuif Cedex, France.

Controlled Clinical Trials 3:227–234 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030227-08$02.75

Table 1 Geographical Breakdown of Trials Registered at the Information Office of the UICC

Geographical area	Number of trials
Africa	25
America	
North America	
Canada	26
Mexico	2
United States	446
South America	
Argentina	6
Brazil	6
Colombia	1
Asia	43
Europe	354
Australia–New Zealand	26
Several areas	10
Total	945

published, and to find out the reasons for such decisions (negative results, for example).

Table 1 shows that 75% of the trials registered were undertaken in the United States and in Europe, where therapeutic improvements are generally reported. Breaking the trials down by tumor sites (Table 2) and comparing these numbers with the incidence rates of these tumors, we see that 25% of the trials were carried out in leukemias and hematosarcomas, which represent only 5%–6% of all malignant tumors, whereas only 10% of the trials were conducted in digestive cancers, which constitute approximately 25% of all malignant tumors.

These observations have to be correlated with the fact that treatments have progressed much more in the field of leukemias and hematosarcomas than in that of the digestive tumors, the two phenomena probably being a con-

Table 2 Distribution by Sites of the 945 Trials Registered at the UICC Information Office, and Related Incidence Rates

Site	Incidence[a]	Percentage of trials by site
G.I. Tract	76.0	10.8
Genitourinary sites	46.9	5.9
Breast	40.4	15.9
Lung	40.0	12.5
Gynecological sites	30.4	6.9
Leukemias and hematosarcomas	23.0	26.7
Miscellaneous	15.3	3.3
Head and neck	15.0	5.4
Brain and nervous system	4.9	4.3
Skin (including melanoma)	4.2	5.4
Bone and soft tissue	2.8	3

[a] Average annual age-adjusted incidence rate per 100,000 population—United States.

sequence of the difference between the efficacy of chemotherapy in these two types of cancer. Variations in the efficacy of chemotherapy contribute for a large part to the differences in the numbers of trials according to sites, as observed in this survey.

Within tumor sites, it is also possible to break down the trials by types of treatment. Table 3 gives an example for head and neck cancer. There is an important disproportion between the treatments, chemotherapy being the most frequently tested. Although all the problems concerning radiotherapy and surgery have not yet been solved, physicians are more reluctant to undertake trials in which the two procedures are compared.

Table 4 gives the characteristics of 27 trials followed through the publication of results. In ten of these trials no statistical difference between the treatments compared could be evidenced. In such cases it is not possible to conclude that the treatments are equivalent. However, since the numbers of patients were large enough, it can be assumed that there were no major differences. Results such as these enable physicians to appreciate the actual value of certain new treatments.

In another ten trials there was a significant difference. In half of these it was in favor of the classic treatment, which is a good argument against the adoption of a new treatment without prior evaluation in an RCT.

In the remaining seven trials, no clear conclusions could be drawn. Such difficulties stem essentially from methodological errors in the design and analysis of the trial, such as wrong timing of the randomization, multiple interim analyses, analyses by subgroups, and so on [2].

Table 3 Types of Treatment Compared in 48 Randomized Trials Registered Between 1968 and 1977 in Head and Neck (UICC)[a]

Treatment	Number of trials
Surgery	
Prophylactic node dissection	1
Versus surgery and Rx	1
	2
Radiotherapy	
With oxygen	3
With calcium perfusion	1
Various NSD	6
Neutrons/photons	2
Prophylactic on nodes	2
Post/Pre-Operative	2
	16
Chemotherapy	
Adjuvant + Rx (pre,concomitant, or post)	12
+ Rx and Surgery	4
Alone (advanced tumor, or comparing regimens)	9
	25
Immunotherapy	
Adjuvant	5

---

[a] Five are abandoned, 10 are published, and 14 are terminated but unpublished.

Table 4 Main Characteristics of 27 Published Randomized Trials (Registered at the UICC Information Office)

Conclusions	Number of trials	Mean number of participating centers	Mean number of patients[a]	Average duration
Treatment(s) under trial at least partially better than classic treatment	5	9	457/346	70 months
Treatment(s) under trial inferior to classic treatment	5	7	217/167	57 months
No difference between treatment(s) under trial and the classic treatment	10	11	322/243	64 months
Uncertain and complex conclusions	7	11	317/256	54 months
Total	27	10	233	69 months

[a] Included/analyzed.

We also can notice that the mean duration of these trials was between 5 and 6 years. This is a major disadvantage since knowledge may progress during this period. As a consequence it may prove difficult to continue running trials for ethical reasons, or the results may already be out of date by the time they are obtained.

Our analysis should be extended to a larger number of trials. Moreover, an attempt should be made to determine the extent to which the results of these trials are being put into practice by the medical community.

SURVEY OF THE STANDARD TREATMENTS IN CANCER

This approach involves many difficulties:

How does one inventory the treatments used at the time of the survey?

How does one define a standard treatment?

How does one ascribe acknowledged therapeutic progress to a given trial?

Needless to say, treatments vary from one country to another, within a given country, and from one physician to another. There is no universal consensus as to what constitutes the best treatment for a given cancer. Should there be one, it would have to be revised regularly to take into account the results of the most recent research.

The reasons for adopting or changing a treatment are multiple, based on individual and collective decisions. Besides clinical research, such elements as backgrounds, clinical experiences, advertising, rumors, available equipment, and medical orientation (surgery, radiotherapy) play a part in the decision-making process. Thus, to determine the modification for which RCTs are wholly or partly responsible is a very difficult task.

Tables 5 and 6 illustrate a survey conducted in our own hospital for head

Table 5 Larynx Treatment Policies[a]

1960–1967		After 1967
Glottic : Surgery ± Rx		T_1, T_2 = Rx N_0, N_1
		T_2, T_3, T_4 = surgery ± Rx N_2, N_3
Supraglottic Total Larynx	Rx 4000 rads ↙ ↘ Rx or surgery ± Rx	T_1, T_2 = Rx
		T_3, T_4 = surgery + Rx BCG adjuvant = randomized trial

Causes of Change

In glottic tumor

Desire to improve the functional result for tumors with good prognosis.

Numerous noncomparative studies showing no difference between series treated either by surgery or by radiotherapy.

In supraglottic tumor

Counteroffensive of the surgeons impressed with the results of a randomized trial conducted on hypopharynx tumors: better results and fewer complications with postoperative Rx [4].

[a] TNM classification from UICC 1969.

Table 6 Oral Cavity Treatment Policies

Before 1970	After 1970
Interstitial curietherapy	
+ Node dissection	– Decrease of use of curietherapy
± Radiotherapy (Rx)	
or	– Increase of use of surgery
Transcutaneous Rx	– Association of transcutaneous radiotherapy and:
Surgery in Few Cases	– Curietherapy
	– I.A. Chemotherapy
Systematic Node Dissection	– Partial suppression of systematic node dissection

Causes of Change

Complications (necrosis) due to curietherapy procedure when performed close to bone, or on large tumor.
Development in the surgical procedures, particularly reconstructive operations.
Results of two randomized clinical trials:
  a. Efficacy of I.A. chemotherapy [5].
  b. Inefficacy of systematic node dissection [6].

and neck tumors, in which we were trying to find an explanation for changes observed between two periods of time. We then extended further this work by studying the published reports of therapeutic results in head and neck cancers. Our intent was to determine for each main site whether or not there was a consensus concerning treatment, which were the reasons leading to the choice of a specific treatment, and what was the degree of correlation between the unsolved problems and the trials being conducted by the various international cooperative groups; similar work has been done in a number of sites other than head and neck by a subgroup established by the UICC Project on Controlled Therapeutic Trials [3].

The first conclusion is that in most cases where there is a consensus, it has been the result of nonrandomized studies, comparative or not. On the other hand, the majority of published trials dealing with chemo- or immunotherapy are randomized. The explanation is that the use of modern statistical methods in the field of clinical trials is much more recent than that of surgery and radiotherapy in the treatment of cancer. As a consequence the main treatment of solid tumor, i.e., surgery and/or radiotherapy, have never been established by RCTs.

RCTs have also been focused on sites with very poor results or in cases where the differences between the treatment procedures were relatively small. It is not surprising, under these conditions, that it is difficult to determine the impact of RCTs, taking into account the fact that many trials have a very narrow impact, limited to the center or to the country where they were conducted. Moreover, the knowledge of a method capable of assessing the efficacy of new therapies is by no means sufficient to make therapy advance. New original therapies must, of course, first be discovered.

Nevertheless it is regrettable that RCTs do not have a stronger impact.

Table 7 Power of Tests for Three Trials Using Adjuvant 5-FU[a]

Trials	Number of patients in trial	Probability of finding a difference ≥0.2
Short-term 5-FU curative surgery	308	0.95
Short-term 5-FU palliative surgery	125	0.68
Long-term 5-FU curative surgery Duke's C	522	0.997
Long-term 5-FU palliative surgery	163	0.80
Long-term 5-FU curative surgery Duke's B + C	189	0.85
Long-term 5-FU palliative surgery Duke's B + C	78	0.49

[a] From ref. 3.

Some of the reasons for this are related to methodological errors. One common error is to consider as a contradiction the fact that no significant difference is evidenced in one of two trials comparing the same treatments whereas there is a significant difference in the other. The absence of a significant difference may be due to the test's lacking power because of an insufficient number of patients (Table 7). When negative results are published, mention should always be made of the power of the comparison.

Other errors have been reviewed in several papers such as inadequate formulation of the problem, repeated interim analyses, and analyses by subgroups.

Clinical trials are often considered as part of the routine activities of physicians and not as genuine research work. Consequently many protocol deviations occur, data are not properly collected, patients are lost to follow-up and physicians are impatient to obtain results. Too many emotional considerations are implicated in the reasoning process of physicians, especially as regards cancer, to permit judgments as objective as those made in experimental research.

Finally, one may consider that the impact of RCTs on treatment protocols for cancer patients has been increasing over the last decade insofar as physicians have become more convinced of the necessity of using adequate methodology to improve the quality and validity of their results. This has made these results more convincing to their colleagues. They permit the rejection of inefficient or toxic methods as quickly as possible and the use of less agressive therapies with equivalent results.

This impact would be increased further if a general strategy determining the most important trials to be undertaken could be defined at an international level. The feasibility of establishing such a strategy for head and neck cancers is being assessed by a subgroup of the UICC Project on Controlled Therapeutic trials.

REFERENCE

1. Flamant R, Fohanno C (Eds): Controlled therapeutic trials in cancer. Geneva: UICC technical report series, vol 32, 1978
2. Armitage P, Bardelli D, Galton DAG, Gehan EA, Higgins GA, Magnus K, Miller AB, Pocock SJ, Saracci R: Methods and impact of CTT in cancer. Geneva: UICC technical report series, Part I, vol 36, 1978
3. Cascinelli N, Davis HL, Flamant R, Keus Y, Lalanne CM, Muggia FM, Rozeneweig M, Staquet MS, Veronesi U: Methods and impact of CTT in cancer. Geneva: UICC technical report series, Part II, vol 59, 1981
4. Vandenbrouck C, Sancho H, Le Fur R, Richard JM, Cachin Y: Results of a randomized clinical trial of preoperative irradiation versus postoperative in treatment of tumors of the hypopharynx. Cancer 39(4):1445–1449, 1977
5. Richard JM, Sancho H, Brugère J, Vandenbrouck C: Intra-arterial methotrexate in head and neck tumors. Europ J Cancer 9:847–851, 1973
6. Vandenbrouck C, Sancho-Garnier H, Chassagne D, Saravane D, Cachin Y, Micheau C: Elective versus therapeutic radical neck dissection in epidermoid carcinoma of the oral cavity. Cancer 46(2):386–390, 1980

Randomized Surgical Clinical Trials for Treatment of Coronary Artery Disease

Lloyd D. Fisher and J. Ward Kennedy

CASS Coordinating Center, Seattle, Washington

ABSTRACT: The impact of randomized clinical trials of surgical therapy for treatment of coronary artery disease is evaluated. Trials are reviewed leading to the following suggestions: 1. Large multicenter trials have a major impact on clinical practice; 2. Small trials have less impact; 3. Findings in agreement with current practice are accepted enthusiastically; 4. Findings disagreeing with current practice are met very critically; and 5. Surgical trials have special problems because of a great variability between clinics, the potential of uneven learning curves, and potential "cross-over" from medical to surgical therapy.

KEY WORDS: *randomized clinical trial, surgical clinical trial, coronary artery disease*

INTRODUCTION

This article considers the impact of randomized clinical trials of surgical therapy for the treatment of coronary artery disease. In particular we are interested in whether the trials had some influence in changing the opinions of experts. We shall also be interested in characteristics of these trials that influence subsequent therapy as opposed to those that had less effect. To the extent possible we shall discuss the impact of trials in a quantitative manner. The data used for that purpose are taken largely from published results of the Coronary Artery Surgery Study (CASS).

We review briefly some of the standard nomenclature for coronary artery disease. The primary diagnostic tool for coronary artery disease is coronary angiography. This is carried out by placing a catheter into the orifice of each coronary artery and injecting a radiopaque contrast material. The information is recorded on x-ray motion pictures (cine angiograms). The images of the coronary arteries are examined for the presence and severity of vessel narrowing or occlusion. There are three major vessels in the coronary artery tree. Disease is often classified in terms of the number of the three major vessels that are diseased. Functionally important narrowing is considered present when a vessel is narrowed more than a fixed percent of its normal diameter.

Received June 14, 1982; revised July 1, 1982.

Address reprint requests to: Dr. Lloyd Fisher, Cass Coordinating Center, University Building, Room 530, 1107 N.E. 45th Street, Seattle, WA 98105, USA.

Controlled Clinical Trials 3:235–258 (1982)

© Elsevier Science Publishing Co., Inc. 1982
0197-2456/82/030235-24$02.75
52 Vanderbilt Ave., New York, New York 10017

Seventy percent, 75%, and 50% have been used as cutoff points for significant disease. In addition to disease of the coronary arteries, the extent of chronic damage to the left ventricular muscle due to prior myocardial infarction (heart attack) is an important factor in determining short-term and long-term survival.

Many individuals come to coronary angiography because of a type of chest pain called angina pectoris. Angina pectoris is caused by lack of adequate oxygen supply to the heart muscle, usually due to narrowed or occluded coronary arteries impeding the blood flow to some portion of the heart muscle. Classical angina pectoris occurs with exercise. Anginal symptoms at rest suggest more severe disease. In addition, an angina pattern rapidly progressing in severity indicates a worsening prognosis. Unstable angina has a variety of definitions. In its most severe form it suggests an impending myocardial infarction.

THE CORONARY ARTERY SURGERY STUDY, CASS, REGISTRY

Published data from the CASS registry are used to illustrate some of the points later in the article. The CASS registry is described in detail elsewhere [1]. At each of 15 institutions all patients studied angiographically for suspected or proven coronary artery disease were enrolled into the CASS registry of 24,959 patients. Of these patients, 780 patients were randomized between coronary artery bypass surgery therapy and medical therapy. As the randomized study of CASS is not yet completed, this portion of the study is not discussed further in this article.

The following is a brief review of a number of randomized trials to evaluate the surgical management of coronary heart disease.

INTERNAL MAMMARY LIGATION

Early surgical trials involved internal mammary ligation. The rationale for ligating, or tying off the internal mammary arteries, was given in a typically enthusiastic article in the lay press [2], entitled "New Surgery for Ailing Hearts":

> The story of this operation begins in 1939 in Genoa, Italy, with the work of Dr. Davide Fieschi, a humble and modest physician. Fieschi, studying the maze of blood vessels that feed the heart, became particularly interested in the two internal mammary arteries—soda-straw-sized, arterial pathways that run down through the center of the chest. He was aware that there was a linkage between the mammaries and the coronary arteries.
>
> If such a linkage existed, couldn't the flow to the heart be increased by tying off the mammaries? Wouldn't the blood, no longer able to flow *downward* through the mammaries, instead back up into the heart, providing a richer supply of nourishment? To Fieschi this seemed 'a simple matter of hydraulics'. As for the area normally fed by the mammaries, blood supply to the chest region is so rich that circulation would be taken up by other blood vessels.

An observational evaluation of the treatment, uncontrolled, by Battezzatti, Tagliaferro, and Cattaneo [3] reported symptomatic improvement in 94.8% of the cases.

The surgical procedure was evaluated using the randomized double-blind technique by two groups. Cobb et al. [4] randomized 17 patients, 8 to ligation and 9 controls. The control group had the incision made in the chest and the artery exposed for ligation. The estimated degree of subjective improvement as given by the subjects was 32% for the ligated patients and 43% for those whose internal mammaries were not ligated. The authors conclude that "bilateral skin incisions in the second intercostal space seem to be at least as effective as internal mammary artery ligation in the therapy of angina pectoris."

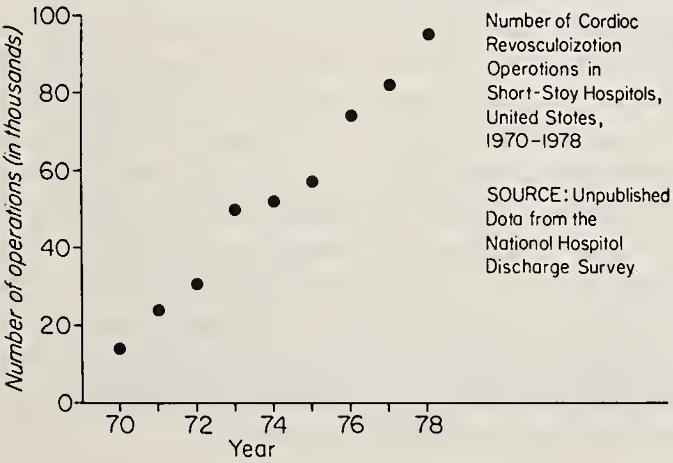
A second randomized study was reported by Dimond, Kittel and Crockett [5]. Eighteen patients with angina pectoris were studied and similar conclusions reached.

The internal mammary artery ligation procedure, although it gained brief popularity at one time, was adopted by only a minority of cardiologists. After the publication of the results of the two trials, the procedure was dropped completely. In the opinion of the authors, this rapid cessation of the procedure was a result of the randomized clinical trials.

CORONARY ARTERY BYPASS SURGERY

Since 1967, the number of bypass operations has increased enormously. Figure 1 shows the number of coronary artery bypass graft operations (CABGs) by year. CABG surgery rapidly became the therapy of choice for individuals with severe angina. The rapid acceptance of this radical new form of therapy stemmed from the striking effectiveness of CABG surgery in the relief of angina pectoris. Today the majority of cardiologists believe that this operation is highly effective in the relief of moderate or severe angina pectoris. (Nevertheless, see Preston [6].) Most of the debate has centered around whether or

Figure 1. Number of coronary bypass operations, in thousands, per year. Estimates are from unpublished data of the National Hospital Discharge Survey.



not the operation prolongs life in specified subsets of individuals. Therefore, the primary thrust of the following discussion shall be with regard to survival.

THE MATHUR AND GUINN RANDOMIZED STUDY OF CORONARY BYPASS SURGERY AND STABLE ANGINA

In August of 1975 Mathur and Guinn [7] reported a randomized study of coronary artery bypass surgery involving 50 patients randomized to surgery and 50 patients randomized to medicine, with follow-up of 8–34 months and a median follow-up of 24 months. At that time period, 6% of the surgical patients and 10% of the nonoperated patients had died. In line with the general observational data, 70% of the surgically treated patients were asymptomatic compared to only 8% of those being treated medically. The authors stated that the quality of life appears to be better in the operated patients during the first 24 months, while the survival data were not statistically significant in either direction. This small study came at a time when bypass surgery was becoming increasingly popular; it is difficult to assess what effect, if any, this study had upon the medical and surgical practice. The sample size was underpowered for the mortality rates involved.

THE PORTLAND VA STUDY OF ACUTE CORONARY INSUFFICIENCY (UNSTABLE ANGINA)

Seldon et al. [8] reported on a randomized clinical trial of 40 unstable angina individuals admitted to the Portland Veterans Administration Hospital between May 1972 and May 1975. In 4 months of follow-up there were no deaths and two myocardial infarctions in the medically assigned and treated patients and one death and three myocardial infarctions in the 21 patients assigned to surgery. The surgical patients had statistically significantly higher functional capacities.

The authors conclude that there is serious doubt about the need for precipitate surgical procedures performed under emergency conditions at irregular hours with the patient accutely ill and experiencing angina at rest. They also note that the medically treated individuals have fewer complications than reported in some prior literature and conclude "the clinical course of our patients treated medically clearly differed from that predicted by published reports, emphasizing the importance of a prospective randomized study to compare in a valid manner different forms of therapy." They also conclude "initial medical management of patients with acute coronary insufficiency followed by elective coronary bypass in patients with continued disabling angina pectoris is a reasonable alternative to emergency bypass." At the time of this study, many believed that acute coronary insufficiency urgently required surgery. The relatively benign course of the individuals randomized to medicine indicated that this was not so. It is difficult for us to assess the impact of the Portland study; it may have prompted the large national cooperative study on unstable angina which followed.

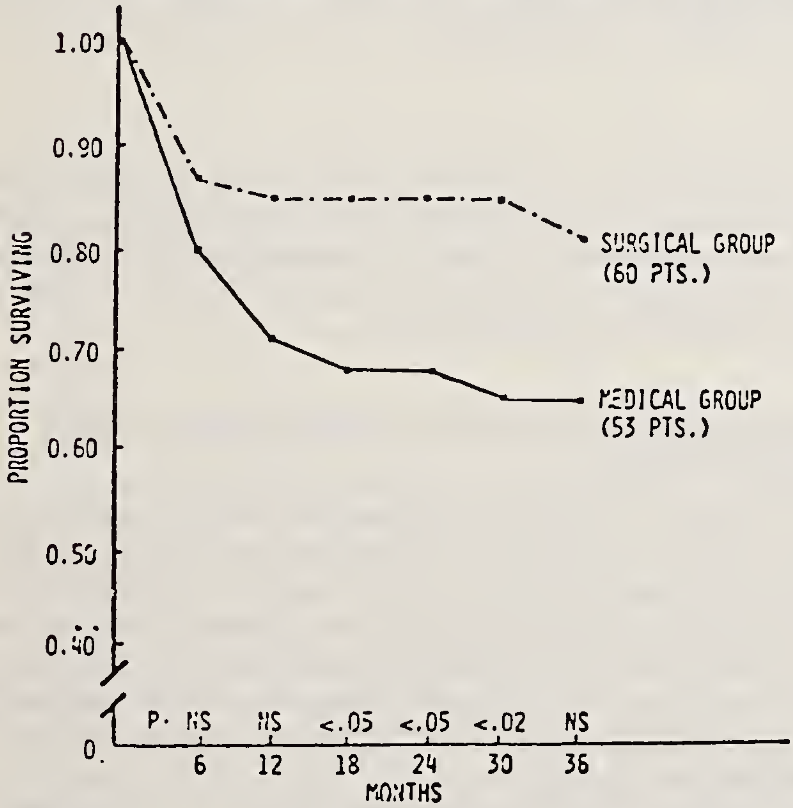
THE VA COOPERATIVE RANDOMIZED STUDY: INDIVIDUALS WITH LEFT MAIN CORONARY ARTERY DISEASE

The Veterans Administration Cooperative Study of Surgery for Coronary Arterial Occlusive Disease has had the most impact among the randomized studies published to date. The VA study began in 1968 as a randomized study of the Vineberg Implant Operation [9]. The Veterans Administration Study shifted to a randomized study of coronary bypass surgery in 1970, with recruitment ending in 1974. Patients were selected to have stable angina pectoris of more than 6 months duration with electrocardiographic evidence of coronary artery disease. To participate in the study the patients had to have a reduction of the luminal diameter in one view of at least 50% in one or more of the major vessels.

The medical–surgical comparison for 113 patients with left main disease was reported in 1975 by Takaro et al. [10]. Figure 2 shows the survival comparison. Surgery was statistically significantly superior to medical therapy in prolonging life.

At the time of the study it was suspected from observational data that left

Figure 2. VA randomized trial of stable angina: Left main patients. Cumulative survival rates of 113 patients with significant lesions of the left main coronary artery by treatment assigned. These patients were randomly allocated to medical or surgical treatment from 1970 through 1974. The numbers in each group refer to the number of patients at zero time. NS: "P" value not statistically significant. Average follow-up is 30 months.



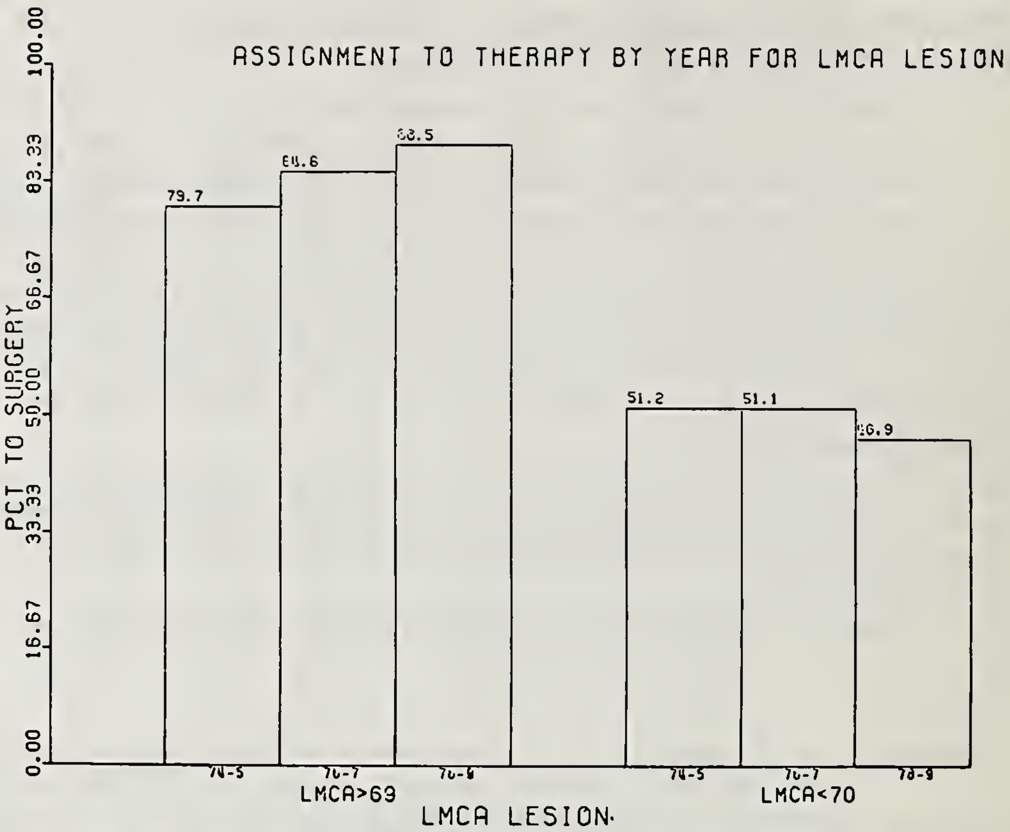


Figure 3. Percent of chest pain patients with and without 70% left main coronary artery disease that receive coronary bypass surgery within 90 days of angiographic study. The patients are from the Coronary Artery Surgery Study (CASS) registry. The percentages are given in three time periods: 1974–1975, 1976–1977, and 1978–1979.

main patients might benefit from surgery. The VA findings were immediately accepted. Figure 3 shows the percentage of chest pain patients with left main disease of 70% or more in the CASS study that received surgery by time [11]. We believe that the increase in surgical management is due, at least in part, to the results of the VA Cooperative Study.

THE GENERAL RESULTS OF THE VA STUDY FOR THE TREATMENT OF CHRONIC STABLE ANGINA

In September of 1977, the Veterans Administration trial reported the initial results of their cooperative study [12] upon the individuals enrolled from January 1972 to December 1974; this was considered a period after the initial development of the angiographic and surgical techniques. The overall surgical mortality was 5.6%. Figure 4 shows the survival data for all 596 patients without left main disease. At 36 months the survival in the medical group was 87% compared to 88% survival in the surgical group, a difference that was not statistically significant.

The article appeared with a lengthy lead editorial by Eugene Braunwald [13]; he made some very cogent points about the importance of the Veterans

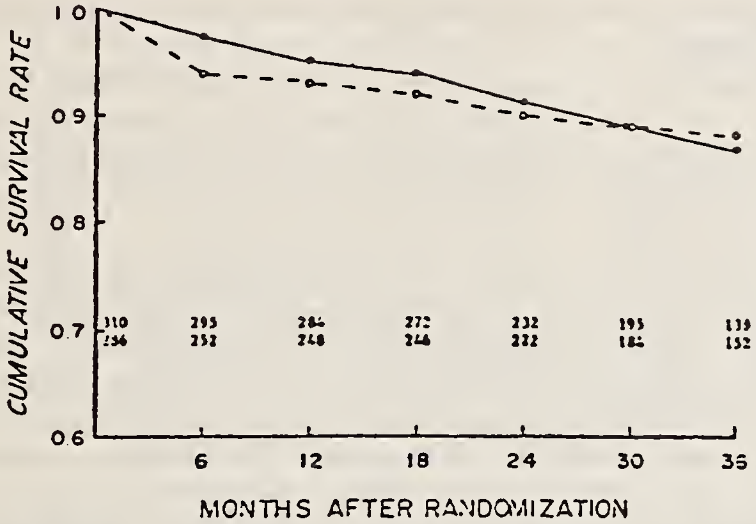


Figure 4. Survival of non-left main patients in the VA stable angina study, 1977. Cumulative survival rates for patients according to treatment assigned for the medical (O) as compared to the surgical (O) group. The numbers of patients at risk in both groups at specific intervals after randomization are indicated below the survival curves (top numbers are medical group, and bottom numbers are surgical group).

Administration study. He noted that "although there is little dispute that severe angina, unresponsive or poorly responsive to medical treatment is an indication for CABG, a principle unresolved question concerns its effect on survival." He points out that although critics of the study will point to an in-hospital mortality of 5.6%, which was excessive at the time the study was carried out, that the excellent medical survival would make it difficult for surgical results to be substantially better. Additionally, the medical survival is better than that reported in the 1960s so that nonconcurrent controls are inappropriate; medical therapy has not stood still. The preliminary results of the Veterans Administration Cooperative Trial are "welcome indeed."

The appearance of the VA results was associated with press conferences; the study received wide, front-page coverage in many papers in the United States.

Reaction of individuals to the VA study was immediate, sustained, and vocal. In the December 29, 1977, *New England Journal of Medicine* an editorial and series of letters appeared. The editorial, by H. H. Hiatt [14], commented on the lessons of the coronary bypass debate, stating that the hotly debated results of the VA trial would be better served if well-controlled studies had been undertaken sooner. Hiatt suggested "the aggregation of the experience of doctors throughout the country, careful analyses of such data, and application of those analyses to medical practice are the exception, rather than the rule, curtailing the usefulness of the medical literature." He went on to say, "professionals, politicians and the public must be educated to understand that advances occur generally in small steps." Finally, he concluded, "some constraints in the spread of new procedures must be accepted. Well designed trials should precede widespread dissemination, as is done to a considerable extent for drugs. Why not for surgical procedures? Why not for costly procedures of all kinds?"

Following this editorial a special correspondence section entitled "A Debate on Coronary Bypass" was presented. The introduction states, "as expected the article. . . . attracted unusual interest . . . Because the question is so important, we are breaking precedent by devoting a special correspondence section to this topic" [15].

Among the points made were the following: the poor results of the study from the surgical point of view may be related to the quality of the surgery [16]. The VA study had poor surgical results (the 5.6% mortality), and fewer grafts than might be desired. The medical results may be poorer than expected [17]. Because of variability, perhaps the analysis should present each of the 13 hospitals' data individually [18]. The study was welcomed as a well-controlled series; reviewers and editors could head off inappropriate conclusions drawn from poorly controlled studies [19]. The operative mortality was too high; the graft patency rate was too low, and the medical mortality is more than twice the annual surgical mortality reported at some centers; there was a high number of "crossovers" [20]. Local experts on health care regulation and planning believe that bypass operations prolong life and many people should have the operation. In fact, with the new results the trial may have saved billions of dollars. Even if the trial had cost as much as 50 million dollars, this was the very high benefit-to-cost ratio [21]. The crossover to surgery for 54 of 310 patients assigned to medicine was too large [22]. In two more years it is likely that surgical survival would be superior and statistically significant [23]. In comparing cost, it is necessary to take into account the cost *after surgery* of the medical and surgical groups in subsequent follow-up. "It is highly regrettable . . . that the facile analysis of Dr. Braunwald should have been quoted in the lay press, thereby falsely reinforcing public misconceptions about expensive and unwarranted surgery" [24]. This study does not consider the appropriate physiologic supporting data for coronary bypass surgery [25]. Patients were too highly selective and the surgical mortality was too high [25].

Dr. Braunwald in his reply mentioned that "thus far, all studies that have attempted to compare the effectiveness of coronary artery-bypass grafts and medical therapy of angina pectoris in concurrently treated patients have concluded that the surgical treatment is quite effective in reducing symptoms but that it does not significantly improve survival, except in patients with obstructions of the left main coronary artery" [27]. The reply of the VA participants [28] among other points mentions that if the three hospitals with the highest operative mortality are excluded, the operative mortality rate is reduced to 3.4%, and there is still no statistically significant difference between the treatment groups.

Several things may be concluded from the publicity, furor, and debate about the VA study. First, large randomized studies are taken extremely seriously. Secondly, findings that confirm prevailing conceptions (such as the left main data) are greeted enthusiastically and implemented. Findings that contradict prevailing conceptions generate much debate and may expect criticism and a fairly rocky road.

Some serious and important considerations arose during the debate: some of these were stated in a letter to the New England Journal from Lawrence I. Bonchek [29]:

> A single operation, coronary artery bypass, has gained unprecedented popularity despite lack of unequivocal evidence that it prolongs the life of most patients. This situation has evoked editorial pleas for prospective randomized studies . . . These proposals reflect how much we have profited from prospective randomized studies of new drugs, but they do not fully recognize the differences between new drugs and new operations.
>
> . . . new operations are introduced tentatively; the indications are uncertain and the risk is high . . . Techniques are refined; mortality, morbidity fall, often precipitously . . . A randomized trial, conducted during such an evolutionary phase, will often be invalidated even before the study's end . . . the Veterans Administration study, has not had the expected impact on coronary surgery because it is being criticized for prematurity, poor operative results and lack of relevance to current experience—not for fundamental flaws in design.
>
> . . . Few standard operations would have passed muster if clinical trial would have preceded general use . . . Even a randomized study begun "late" can be rendered obsolete by advances in surgical technology.
>
> We should resist the almost religious fervor of those who have sanctified randomized studies as the only means of learning the truth—if scientific journals and program committees had always rejected reports on non-randomized trials, as some say they should, we would not have the benefits of appendectomy for appendicitis, pylorotomy for pyloric stenosis, or countless other standard operations . . .
>
> A system of intrinsic controls in the practice of surgery already exists. It is based on the referral of patients for surgical therapy by primary physicians, who are not surgeons and are rewarded only by good therapeautic results.[1]

Further reaction to the VA study was presented in JAMA [30].

Participants in the VA study responded, at least in part, to criticism of the study [31–34] and replied to some of the points made by antagonists. Some of the VA points [31] are made in the following Figures 5 and 6. Figures 5a–5c show that the one- , two- , and three-vessel surgical survival is comparable to other existing surgical series during the time period of the VA study. Figures 6a–6c show that the medical survival for one- , two- , and three-vessel disease is better than would have been expected from the extant literature.

We conclude that the Veterans Administration study did have a strong effect on the experts in the field, convincing some that in fact the effect of coronary bypass operation on survival was not proven except in those patients with left main coronary artery disease. The study changed the tenor of the bypass debate to one requiring more justification when survival is the primary motive for bypass surgery (as opposed to relief of anginal symptoms).

At a recent consensus conference [35], additional data from the VA study showed that in three-vessel disease (without left main disease) there was a statistically significant difference in favor of surgery when considering the 10 of the 13 sites with lower mortality.

---

[1] Reprinted by permission from the *New England Journal of Medicine*, **301** pp. 44–45, 1979.

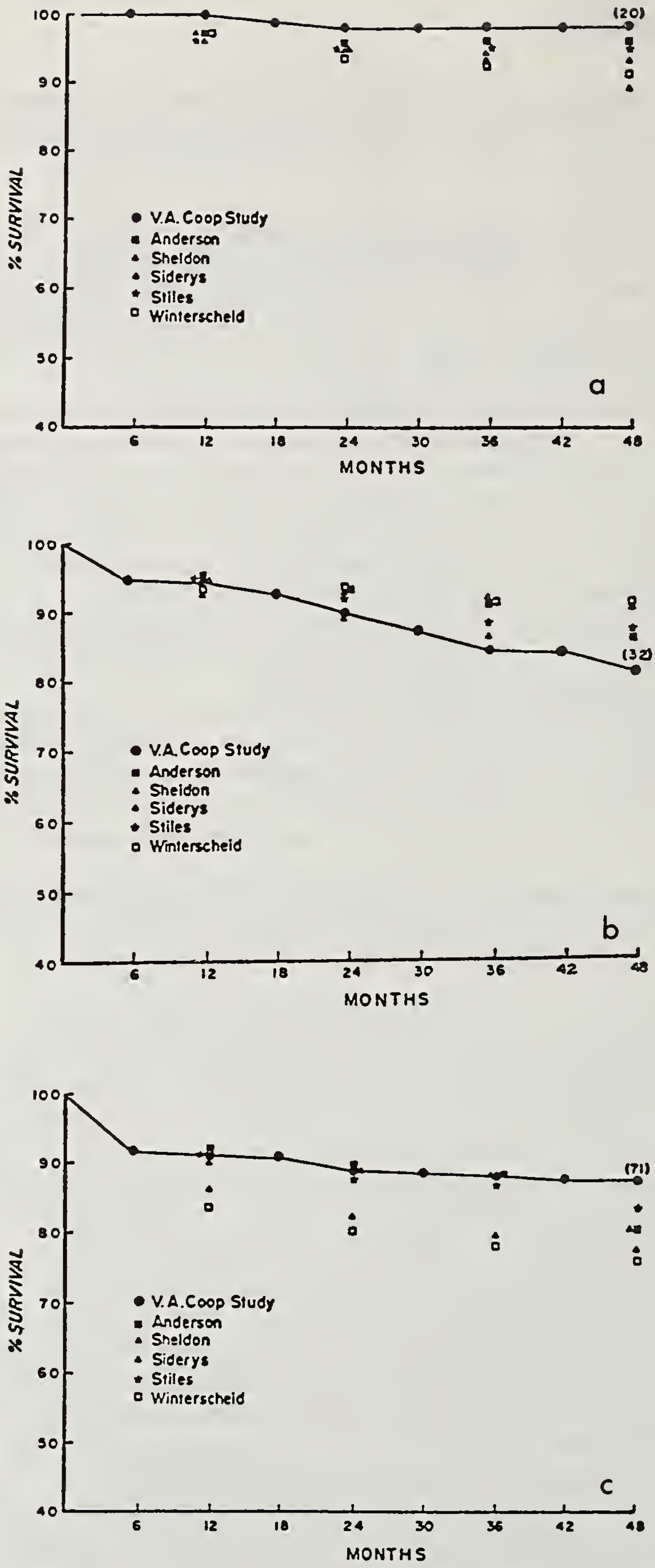


Figure 5. VA Stable Angina Study surgical survival compared to other contemporary studies. (a) One-vessel disease; (b) two-vessel disease; (c) three-vessel disease.

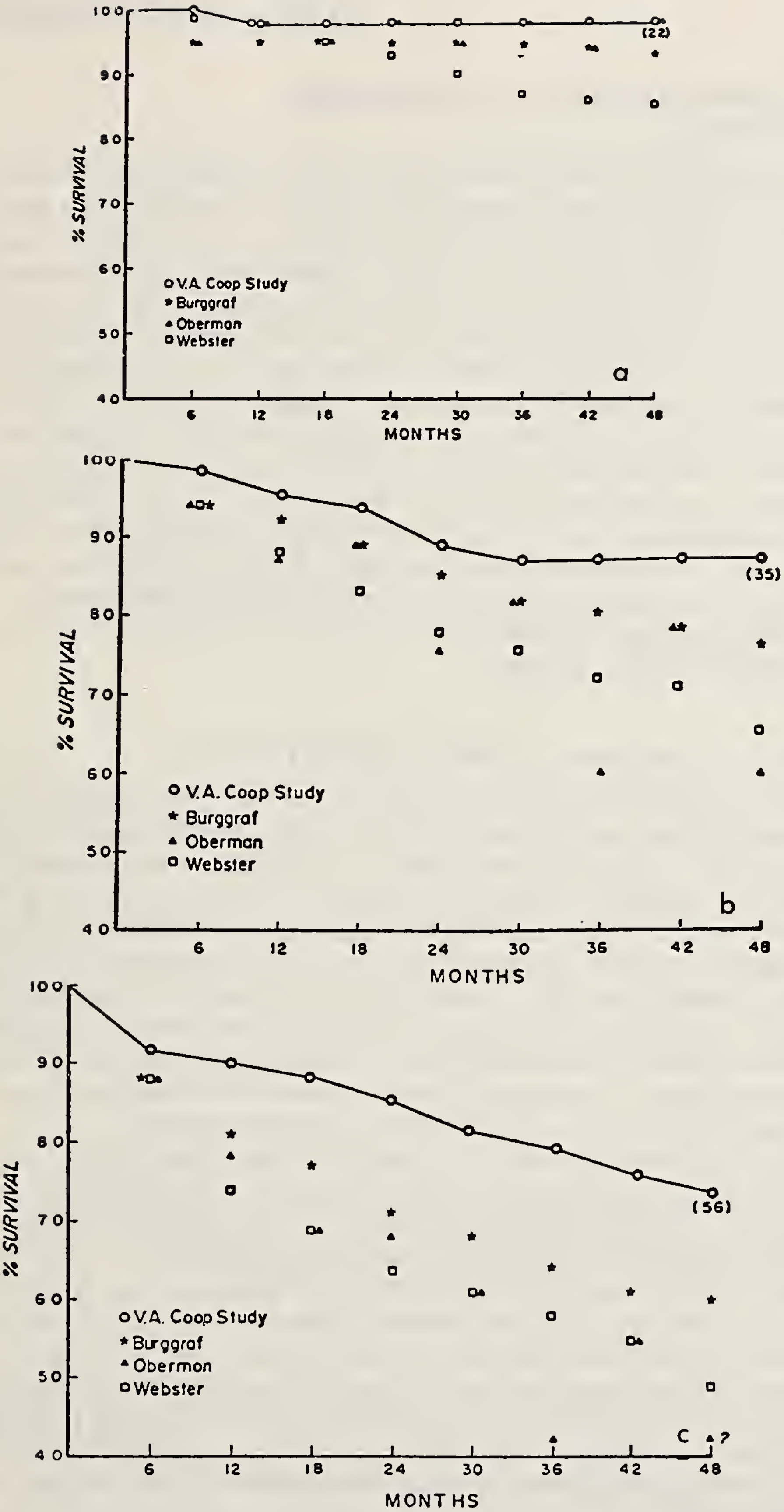


Figure 6. VA Stable Angina Study medical survival compared to other contemporary studies. (a) One-vessel disease; (b) two-vessel disease; (c) three-vessel disease.

NATIONAL COOPERATIVE STUDY OF UNSTABLE ANGINA PECTORIS

Nine cooperative medical centers under the sponsorship of the National Heart, Lung, and Blood Institute between 1972 and 1976 randomized a total of 288 unstable angina patients between immediate coronary bypass surgery and medical therapy [36,37]. The in-hospital mortality was 5% in the surgical group and 3% in the medical group. The rate of myocardial infarction during the last 3 years did not differ statistically between the surgical and medical groups. During an average follow-up period of 30 months, 36% of the medically treated patients had coronary bypass surgery to relieve unacceptable angina. The authors conclude that "The results indicate that patients with unstable angina pectoris can be managed acutely with intensive medical therapy . . . Later, elective surgery can be performed with a low risk and good clinical results if the patients angina fails to respond to intensive medical therapy" [37]. It is the opinion of the authors that this study did have an effect on practice. It has resulted in less urgent and emergent surgery in these patients.

A follow-up paper [38] looked at the patients with ST elevation during pain with roughly similar findings.

THE EUROPEAN CORONARY ARTERY SURGERY STUDY

This European randomized clinical trial of coronary artery bypass surgery in stable angina pectoris was performed on 768 men, less than age 65 with angina pectoris and at least 50% obstruction in two or three major vessels and good ventricular function. The trial reported two-year survival results [39] and actuarial results up to five years [40]. The operative mortality was 3.6% in all the operated patients and in the last third of the enrollment period the operative mortality was 1.5%. Symptomatic improvement was significantly better in the surgery cases. At the time of the first report there was no statistically significant difference between the two treatments among those patients with left main disease or two-vessel disease. Three-vessel cases had a statistically significant difference in favor of improved surgical survival.

In the second paper (with the survival curves given in Figs. 7a–7d) the survival of individuals randomized to surgery was significantly better for the total group, with 5-year survival of 84.1% for the medical group and 93.5% for the surgical group. There was a significant difference ($p = 0.037$) in those patients with left main disease (61.7% vs. 92.9%). The three-vessel survival is also statistically significantly different (84.8% vs. 94.9%). There was not a statistically significant difference ($p = 0.668$) among the cases with two-vessel disease. Of the 373 medically randomized patients, 69 subsequently received coronary bypass surgery. The European study group correctly points out that their study compares a policy of immediate bypass surgery with a policy of no bypass surgery unless specifically indicated by further changes in symptomatic status.

The reception of the European study has been good. The study has not resulted in the type of national debate generated by the Veterans Administration Study but is often cited and discussed as justification for coronary bypass surgery.

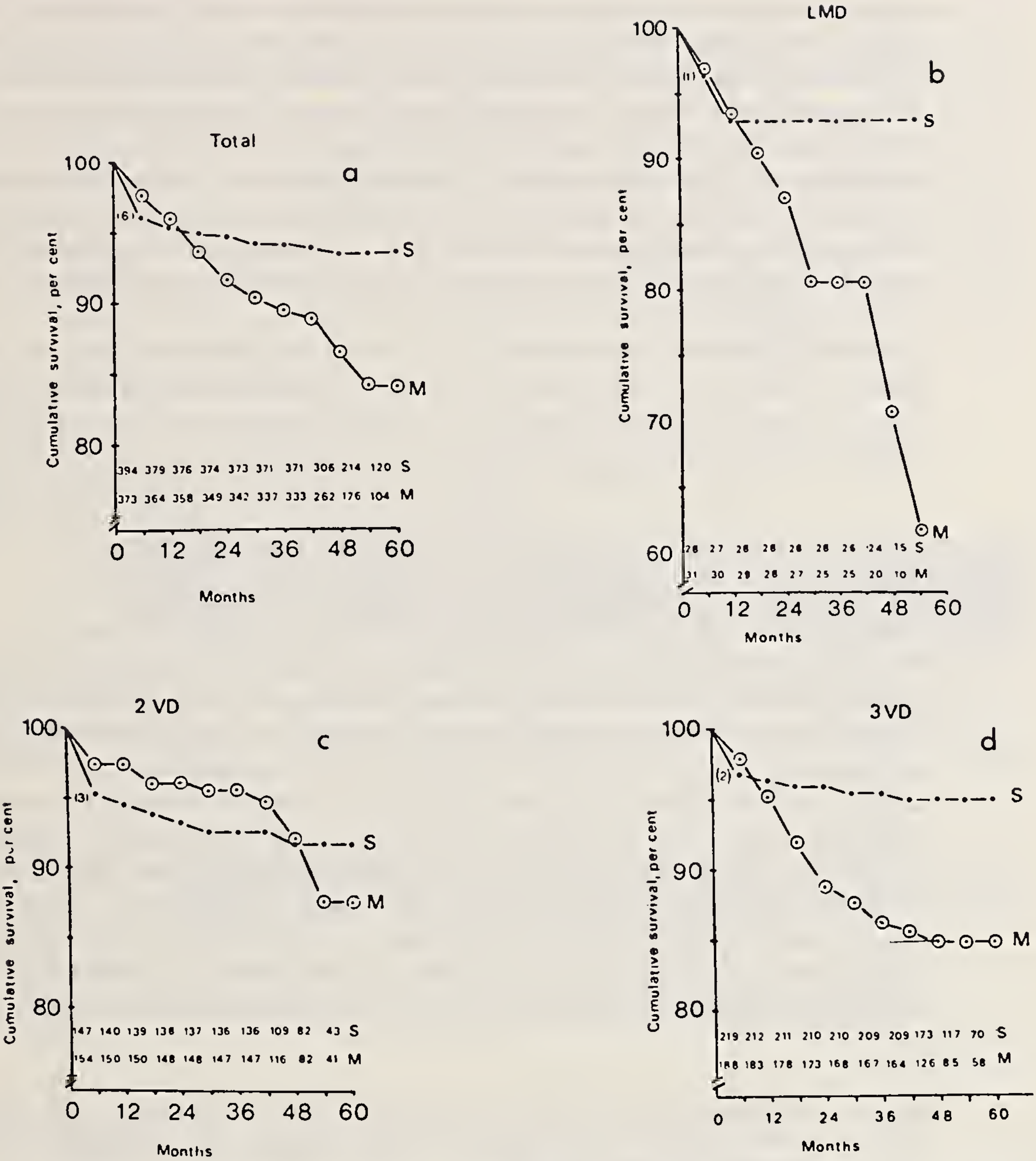


Figure 7. European Coronary Surgery Study Group survival results. (a) Medical versus surgical survival for the entire population. (b)–(d) left main, two-vessel disease, and three-vessel disease results, respectively. Note the interrupted vertical scales.

THE OREGON STUDY OF CORONARY BYPASS SURGERY FOR STABLE ANGINA

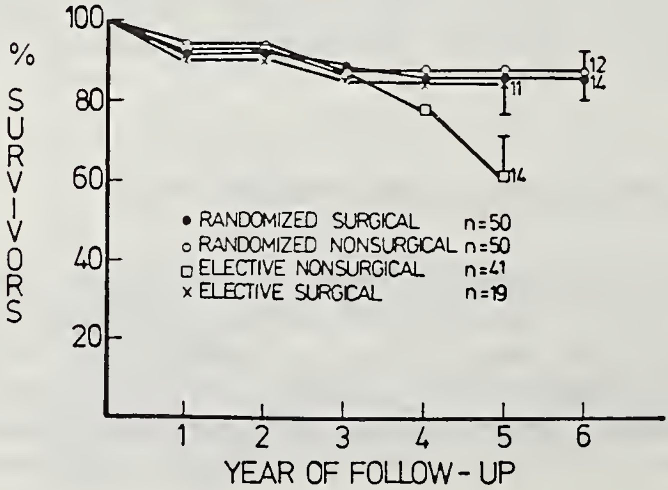
A randomized study for stable, disabling angina was carried out in Portland, Oregon by Kloster et al. [41]. There were 100 patients randomized, 49 to the medical group and 51 to the surgical group. A criterion of 50% or greater reduction in the luminal diameter was considered noteworthy coronary disease and used in the study. Forty-six percent had three-vessel disease, 36% had two-vessel disease, and 18% had one-vessel disease. There were five deaths in those individuals randomized to medicine and four in those individuals randomized to surgery. There were myocardial infarctions in 8 of the individuals randomized to medical treatment and 10 individuals randomized to surgery. There was no statistically significant difference in the survival of the two groups, but considering "terminating events," patients with three-vessel disease who were treated surgically had fewer events. "Terminating events" included death, new myocardial infarction, and unstable angina pectoris persisting despite conventional medical therapy. As in other studies, individuals randomized to surgical therapy had more symptomatic improvement ($p < 0.01$).

Although the study is accepted and quoted when reviewing the controlled studies of angina it clearly suffers from considerably less statistical power than the larger Veterans Administration and European randomized studies. Overall, it has certainly had less impact and is less often cited in the literature.

THE NEW ZEALAND STUDY OF SURGICAL AND NON-SURGICAL MANAGEMENT FOR ASYMPTOMATIC PATIENTS WHO HAVE RECOVERED FROM A RECURRENT MYOCARDIAL INFARCTION

The study group in Aukland, New Zealand, reported on a randomized clinical trial of 100 individuals who had suffered two or more myocardial infarctions [42,43]. The patients were asymptomatic and did not have what was considered quite poor left ventricular function. The term asymptomatic was also used to include those who had angina that was not considered severely disabling. Individuals considered for the study but given elective surgical or nonsurgical treatment were also followed. Figure 8 presents the survival curves for those randomized and also those with elective, therapeutic assignment. There is no statistically significant difference in the survival in the individuals randomized to medical and surgical therapy. The authors comment that "The striking finding in the study is not so much the lack of any apparent difference

Figure 8. Results of the New Zealand randomized trial of "Asymptomatic" patients with advanced coronary artery disease. Results for individuals not randomized are also reported.



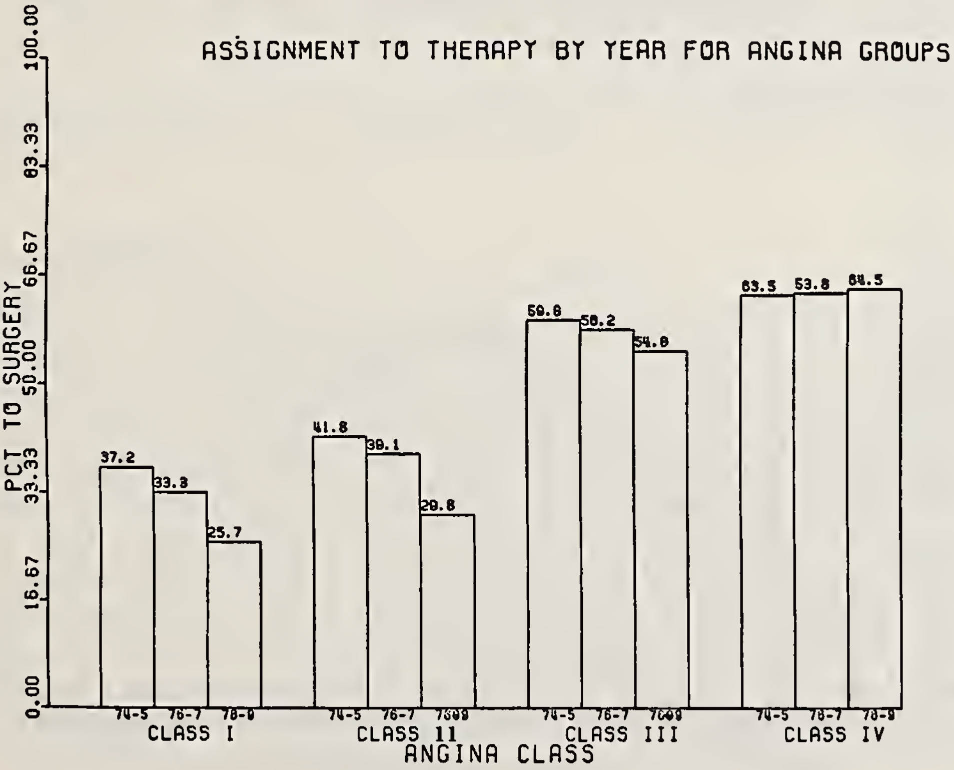
in mortality between surgical and non-surgical patients, as the unexpectedly low mortality rate in the investigative patients regardless of treatment. The annual mortality of 3–4% per year contrasts with the annual mortality rate of 10–15% per year, for patients with disease of equivalent severity." They also note that "the low mortality rate in the present study does agree, however, with that found in the Veterans Administration trial reported in 1977 and it supports a trend towards the lower mortality rate from ischemic heart disease during the last 10 years." They finally conclude, "Our results did not give strong support for the use of surgery in relatively asymptomatic patients with 3-vessel disease, who have survived more than one infarction."

This small study has not received wide discussion within the United States cardiologic community.

POSSIBLE IMPACT OF SURGICAL RANDOMIZED TRIALS AS EVALUATED FROM THE CASS REGISTRY

The participants in the Coronary Artery Surgery Study are blinded to the end-point data of the CASS randomized trial. For this reason the assignment to patient therapy of the nonrandomized patients is evaluated on the basis of evidence external to the CASS trial. A CASS study group evaluated the reasons people were assigned to therapy [11,44]. To identify cases that were elective surgery, surgery within 90 days was used. It was felt that the large majority of surgery after that time period would be due to changing clinical

Figure 9. Percent of patients receiving coronary artery bypass surgery by Canadian Heart Association anginal class and time period in the Coronary Artery Surgery Study.

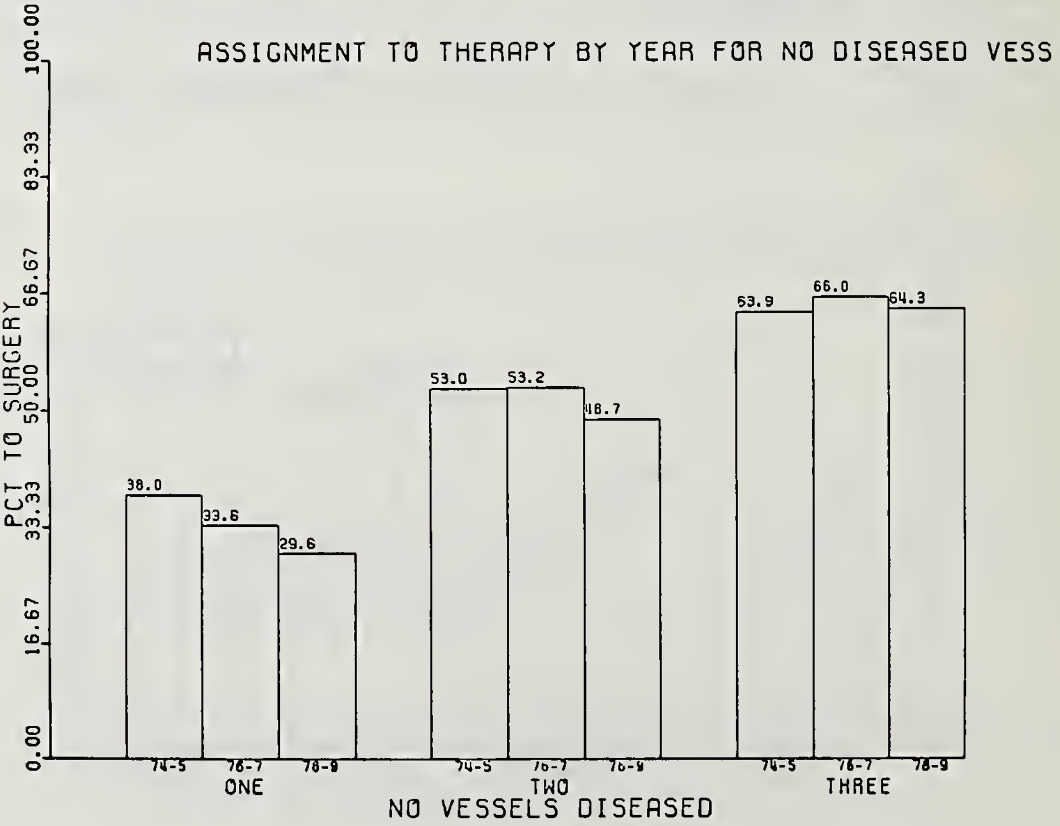


patterns after the base-line evaluation. Patients studied all had chest pain as a syndrome; the patients also had at least one-vessel disease and at least one vessel considered operable.

Figure 9 shows the experience with time in this CASS registry study for individuals with anginal chest pain related to exertion. The Canadian Heart Association anginal classification ranges from I (mild) to IV (very severe). We see that the amount of surgery for individuals in Class IV, who have extreme disability, has remained essentially constant, whereas in Class III, there may have been a slight decrease. In each of the anginal Classes I and II the percent receiving surgery has decreased. Lower anginal classes tend to be associated more often with less severe disease. Thus, although all of the individuals considered in this figure were candidates for surgical therapy, having at least 70% stenosis in one or more vessels and at least one operable vessel, this may reflect in part reactions to the randomized trials which have not shown that survival is prolonged in individuals with one or two vessels diseased.

Figure 10 shows the percentage of individuals going to surgery within 90 days for one- , two- , and three-vessel disease. There has not been an increase in surgery for three-vessel disease that could be attributed to the report of the European randomized trial. In patients with one-vessel disease there has been a steady drop in the percent receiving surgical therapy over the enrollment period of CASS. It is the opinion of the authors that this is the result of a combination of the results of the Veterans Administration Randomized

Figure 10. Percent of patients receiving coronary artery bypass surgery by number of vessels diseased and time period in the Coronary Artery Surgery Study.



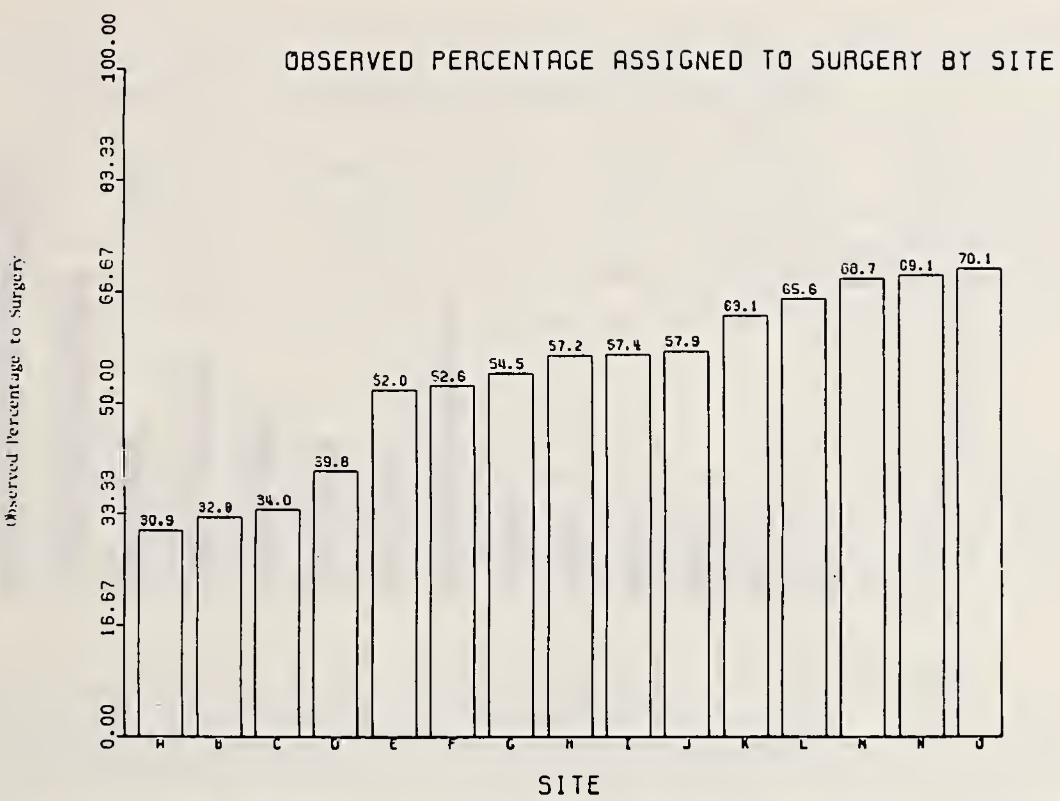


Figure 11. Percent of patients receiving coronary artery bypass surgery by clinical site within the Coronary Artery Surgery Study.

Study and the other randomized studies as well as an increasing appreciation of excellent medical survival in patients with 1-vessel disease.

There is a considerable difference in surgical philosophy between clinics in this country. Figure 11 presents the percent of patients as described above receiving surgery at the 15 clinical sites of the Coronary Artery Surgery Study. There are differences in the patient population, but such characteristics cannot adjust for most of the variability observed.

VARIABILITY IN SURGICAL MORTALITY IN CASS

Preparatory to the discussion of this article it will be useful to consider the variability in surgical mortality rates by clinic. Figure 12 presents the operative mortality figures for CABG only patients at the CASS sites. The white bars in the scale on the left give the range of operative mortality rates. In CASS the operative mortality varied from 0.3% to 6.4%. The mean operative mortality for all the sites was 2.3% [45]. Using a multivariate estimated risk [45], an observed-to-expected mortality ratio can be computed. This is given by the striped bars with the scale on the right. There is still great variability in performance even when one takes into account the characteristics of the patient population. When adjusting for the characteristics of the patient population, sites 8 and 14 perform at approximately the same level, although the observed operative mortality is over twice as high in site 14. Thus, one cannot judge the importance of operative mortality without taking into account the

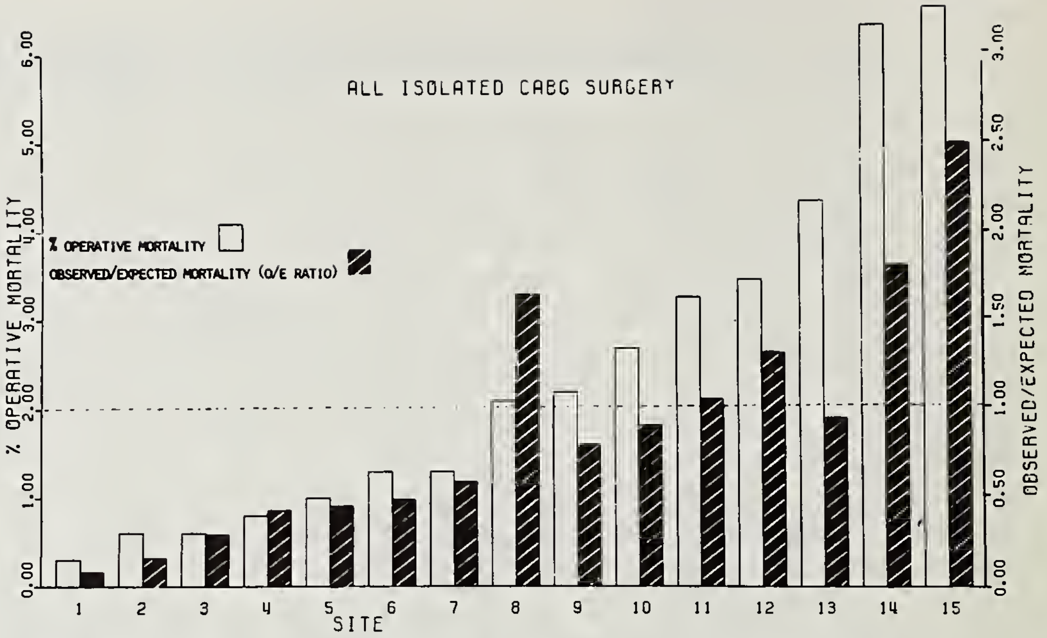
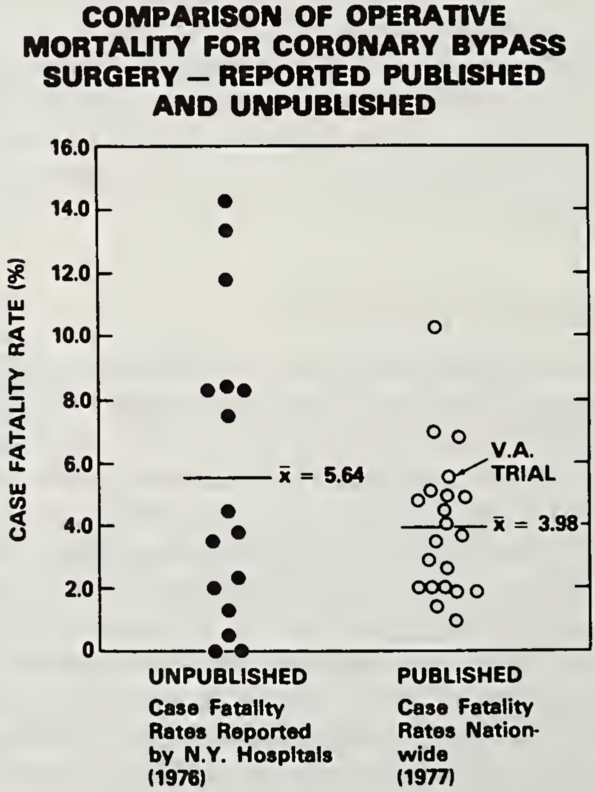


Figure 12. Operative mortality for isolated coronary artery bypass surgery in the Coronary Artery Surgery Study. The open bars give the percent of operative mortality by CASS site. The striped bars give, by site, the observed to expected mortality rates (see text for a more detailed explanation.)

Figure 13. Some published and unpublished surgical mortality rates. Prepared by Thomas Chalmers, M.D.


COMPARISON OF OPERATIVE MORTALITY FOR CORONARY BYPASS SURGERY — REPORTED PUBLISHED AND UNPUBLISHED

patients that compose the surgical series. Nevertheless, even after adjusting for these factors, there was a marked, highly statistically significant difference in the operative mortality rates between the CASS clinics.

This same point can be made in terms of published and unpublished results. Figure 13, prepared by Thomas Chalmers, shows on the right mortality rates from published papers in 1977. On the left is the case fatality rates reported by New York hospitals which were required by law to report their mortality. Of great interest to the future discussion is the wide varibility reported in case fatality rates in each of these two situations.

DISCUSSION

The discussion is organized around questions that were the aim of the conference: the impact of clinical trials on experts and the characteristics that lead to advances. We also consider some problems that are more severe in conducting surgical clinical trials than most other clinical trials.

IMPACT OF RANDOMIZED TRIALS UPON EXPERTS

It is difficult to define who is an expert on the appropriate use of coronary bypass surgery for individuals with coronary artery disease. We shall arbitrarily think of experts as those who are significantly involved in cardiovascular research and/or publish in cardiology and/or thoracic surgical journals.

From our review of the impact of randomized trials, it appears that large multi-institutional randomized trials are taken more seriously than smaller, single-institution trials. This may reflect an appreciation on the part of the scientific community that trials of chronic disease need larger numbers of patients than can be generated at one clinical site. The large reaction and scientific debate that resulted from, and to some extent still surrounds, the Veterans Administration randomized trial of coronary bypass surgery shows the importance of this type of investigation. The negative findings of smaller, single-institution trials have, by comparison, generated relatively little discussion. The VA, European, and Unstable Angina Randomized Clinical Trials are often cited by research-oriented individuals. We feel this reflects, in part, an appreciation of the difficulties of obtaining valid comparisons and a widespread although by no means unanimous consensus that randomized clinical trials do present a stronger scientific inference.

We also suggest that for immediate acceptance of the results of a trial, it is not enough that the trial be well performed according to current standards of medical research. It is our opinion that the Veterans Administration was well run and adhered to good clinical research patterns with a surgical mortality that was acceptable *at the time*. The debate arose when surgical therapy was not shown to be favorable in some subsets of patients. The European trial with a statistically significant difference in favor of surgery was widely accepted; when findings oppose generally accepted views, one can expect a large amount of criticism. This is not unexpected; presumably past opinions are founded upon some facts (perhaps very weak and observational, but nevertheless data). In addition there is the natural human reluctance to accept change. Also, with wide variability in surgical mortality (and possibly long term outcome) there may be criticism from clinics that validly do a better job.

Perhaps it is fair to assess expert views as those resulting from a recent consensus exercise on coronary bypass surgery sponsored by the National Heart, Lung and Blood Institute. Randomized studies were cited prominently in their position paper [46].

CHARACTERISTICS THAT LEAD TO ADVANCES

In order to decide what characteristics of clinical research lead to advances, it is necessary to decide what is an advance. For the purposes of this discussion we define an advance as a shift in opinion that leads to more effective treatment of individual patients.

It should *not* be inferred that advances arise only from randomized clinical trials. Randomized clinical trials began in the mid 1940s; there were certainly considerable advances in medical knowledge and treatment before that time. Advances can be made with observational studies. The problem is one of unappreciated statistical noise. One has less confidence in the results of observational data analyses; in order to make true advances observational analyses may take longer and more individuals may receive an inferior therapy. Appropriate use of randomized clinical trials can lead to swifter and more accurate evaluation of therapy, and consequently lead to better overall health care.

However, there are significant problems involved in mounting randomized clinical trials in chronic diseases that have low short-term mortality. The number of patients needed is large because of the relative rarity of the endpoints. Even the large trials of coronary artery bypass surgery were small when one considers the statistical power of these trials. Although (as Dr. Paul Meier remarked with regard to the potential study of percutaneous transluminal coronary angioplasty [47]) "if we don't do the trial, what will we do instead?" It is probably better to get a smaller amount of quality information than none at all. Nevertheless the general scientific and public appreciation of trials is that they come to one final conclusion.

The most fundamental work is the development of a therapeutic modality. Once a therapeutic modality is developed the characteristics of a trial that lead to advances are randomized trials large enough to have at least moderate statistical power for detecting clearcut differences in therapy.

PROBLEMS MORE SEVERE IN SURGICAL CLINICAL TRIALS

As mentioned above, there is a definite learning curve in the development of new surgical treatments. Dr. Bonchek argues that many beneficial surgical therapies would never have been adopted had they been required, from the first patient, to undergo randomized clinical evaluation. What he says may be true. This does, however, raise ethical problems about the tradeoffs between the individual good and the long-term good of future patients. We believe individuals have the right to accept a reasonable risk for societal good. Nevertheless, if it is generally accepted that new surgical therapies usually have poor early results, it is difficult to know how these therapies can best be introduced into the management of patients.

There is also a statistical problem for clinical trials to take into account a learning curve. Adequate statistical consideration and evaluation of such questions have not yet taken place. It is difficult to imagine any such procedure that would not be based upon some assumptions about the learning process if it is there.

Thomas A. Preston [48], in commenting on an article by Takaro and the VA Study Group [34], mentions that "a more serious criticism generally not voiced is the validity of extrapolation from then to now. If the results were valid for the study period of 1972 to 1974, it is less certain that five years later the same comparison is valid." Of course the purpose of any trial is to find the value in current and future patients. There is no entirely satisfactory answer.

We disagree with Dr. Bonchek that the referral system is a built-in control of adequate quality. Nevertheless, controlled research is essential for surgical techniques that are being applied to large numbers of patients. If the medical or surgical therapy changes over time, then ongoing evaluation is necessary; this is especially true when dramatic changes in medical therapy may also be occurring.

Problems are posed by the wide variability in surgical results between clinics. It is not easy to assemble an excellent surgical team. The results depend upon much more than the individual surgeon. The anesthetist's role is extremely important; the nurse's role and the pre- and postoperative care all enter into the overall results. The considerable variability in the CASS study represents variability among sites that were selected by competitive peer review to have (among other things) high-quality thoracic surgery programs. One would suppose that there is a much wider variability on the national level. In what sort of surgical sites does one want to compare surgical to medical therapy? Should it be a group "representative" of the care that is delivered in the nation as a whole? Should it be representative of "the best" care available, even when such care cannot reasonably be expected to be delivered in many or even the majority of institutions in the country? We believe that such research should be done in the country's better clinics. It may be unfortunate that the health care system is not set up to react formally to such differences except through adverse reactions to extremely poor performance. Although the referral system may tend to channel patients to institutions with better care, it is not structured to perform this task efficiently.

The "crossover" problem is particularly severe in some surgical trials. For coronary bypass surgery there is the crossover to surgery among individuals randomized to medical therapy for pain relief. We feel that the term "crossover" is a misnomer in this context. Because of the widespread agreement that coronary artery bypass surgery is an effective therapy for relief of severe angina, there is a necessity to offer coronary bypass surgery for severe angina which is under poor medical control. Thus, current trials really compare two policies: one recommending immediate bypass surgery versus the second policy of treating with intensive medical therapy and reserving coronary bypass surgery for individuals whose pain is intractable to medical therapy or whose condition changes.

CONCLUSION

From this review we conclude the following:

1. Large multicenter clinical trials of CABG surgery have had a major impact on clinical practice.
2. Smaller trials have had a much lesser impact on clinical practice.
3. Findings in agreement with current medical opinion will be adopted enthusiastically.
4. Findings from large randomized trials in disagreement with current medical opinion will be very critically reviewed.
5. Trials of surgical therapy present considerable challenge due to
    a. great variability in results between clinics;
    b. the possibility of a substantial learning curve, unevenly spread over time; and
    c. if the surgery is being evaluated for longevity, but is also used for pain relief, "crossovers" from medical to surgical therapy may be substantial.

REFERENCES

1. Principal Investigators of CASS and Their Associates: National Heart, Lung and Blood Institute Coronary Artery Surgery Study. Circulation, 63, Part II, American Heart Association Monograph, number 79, 1981
2. Ratcliff JD: New surgery for ailing hearts. Reader's Digest 71:70–73, 1957
3. Battezzati M, Tagliaferro A, Cattaneo AD: Clinical evaluation of bilateral internal mammary artery ligation as treatment of coronary heart disease. Am J Cardiology 4:180–183, 1959
4. Cobb LA, Thomas GI, Dillard DH: Merendino KA, Bruce RA: An evaluation of internal–mammary–artery ligation by a double-blind technique. New Engl J Med 260:1115–1118, 1959
5. Dimond EG, Kittle CF, Crockett JE: Comparison of Internal Mammary Artery Ligation and Sham Operation for Angina Pectoris. Am J Cardiology 5:483–486, 1960
6. Preston TA: *Coronary Artery Surgery: A Critical Review* New York: Raven Press, 1977
7. Mathur VS, Guinn GA: Prospective randomized study of coronary bypass surgery in stable angina. The first 100 patients. Circulation Suppl 1, 51, 52: I-133–I-140, 1975
8. Seldon R, Neill WA, Ritzmann LW, Okies JE, Anderson RP: Medical versus Surgical Therapy for Acute Coronary Insufficiency. A Randomized Study. New England J Med 293:1329–1333, 1975
9. Veteran's Administration Cooperative Group for the Study of Surgery for Coronary Arterial Occlusive Disease, report prepared by Detre K, Hultgren H, Takaro T: Veteran's Administration Cooperative Study of surgery for coronary arterial occlusive disease. III. Methods and baseline characteristics, including experience with medical treatment. Am J Cardiology 40:212–225, 1977
10. Takaro T, Hultgren HN, Lipton MJ, Detre KM, Participants in the Study Group: The Veteran's Administration Cooperative Randomized Study of surgery for coronary arterial occlusive disease. II. Subgroup with significant left main lesions. Circulation Suppl 3, 54:III-107–III-117, 1976
11. Kennedy JW, Killip T, Fisher LD, Alderman EL, Gillespie MJ, Mock MB: The clinical spectrum of coronary artery disease and its surgical and medical manage-

ment. 1974–1979 from the Coronary Artery Surgery Study (CASS). In: The Proceedings of the Consensus Meeting of Economic, Social, and Ethical issues for Coronary Bypass Surgery (to be published).

12. Murphy ML, Hultgren HN, Detre K, Thomsen J, Takaro T, Participants of the Veteran's Administration Cooperative Study: Treatment of chronic stable angina. A preliminary report of survival data of the randomized Veteran's Administration Cooperative Study. New Engl J Med 297:621–627, 1977
13. Braunwald E: Coronary artery surgery at the crossroads. Editorial. New Engl J Med 297:661–663
14. Hiatt HH: Lesions of the coronary-bypass debate. Editorial. New England J Med 297:1462–1464, 1977
15. Relman AS: Introductory comments to special correspondence. A debate on coronary bypass. 297:1464, 1977
16. Laurie GM, Morris GC Jr, Howell JF, Chapman DW: Letter. New Engl J Med 297:1464, 1977
17. Collins JJ Jr, Cohn LH, Koster JK Jr, Mee RBB: Letter. New Engl J Med 297:1464–1465, 1977
18. Towne WD: Letter. New Engl J Med 297:1465, 1977
19. Spodick DH: Letter. New Engl J Med 297:1465–1466, 1977
20. Bonchek LJ, Brooks HL: Letter. New Engl J Med 297:1466–1467, 1977
21. Newhauser D: Letter. New Engl J Med 297:1467, 1977
22. Guss SB: Letter. New Engl J Med 297:1467, 1977
23. Greene DG: Letter. New Engl J Med 297:1467–1468, 1977
24. Maggs PR: Letter. New Engl J Med 297:1468, 1977
25. Rowe GC: Letter. New Engl J Med 297:1468, 1977
26. King SB III, Hurst JW, Logue RB: Letter. New Engl J Med 297:1468–1469, 1977
27. Braunwald E: Letter. New Engl J Med 297:1469–1470, 1977
28. Murphy ML, Hultgren HN, Detre K, Thomsen J, Takaro T: Letter. New Engl J Med 297:1470, 1977
29. Bonchek LI: Sounding board. Are randomized trials appropriate for evaluating new operations? New Engl J Med 301:44–45, 1979
30. Annexton M: Medical news: Vascular surgeons rebut coronary surgery criticism. J Am Med Assoc 240:1218, 1978
31. Read RC, Murphy ML, Hultgren HN, Takaro T: Survival of men treated for chronic stable angina pectoris. A cooperative randomized study (with discussion). J Thoracic Cardiovas Surg 75:1–16, 1978
32. Hultgren HN, Takaro T, Detre KM, Murphy ML: Evaluation of the Efficacy of Coronary Bypass Surgery—I. Am J Cardiology 42:157–160, 1978
33. Hultgren HN, Takaro T, Detre KM, Murphy ML: Commentary. Aortacoronary–artery–bypass assessment after 13 years. J Am Med Assoc 240:1353–1354, 1978
34. Takaro T, the Participants in the Veteran's Administration Cooperative Study of Surgery for Coronary Arterial Occlusive Disease: Results of a randomized study of medical and surgical management of angina pectoris. World J Surg 2:797–809, 1978
35. Takaro T, Hultgren HN, Detre KM, Peduzzi P: The Veteran's Administration Cooperative Study of unstable angina: Current status. *The Proceedings of the Consensus Development Conference in Coronary Artery Surgery*, 1981 (in press)
36. Unstable Angina Pectoris Study Group: Unstable angina pectoris: National Cooperative Study Group to compare medical and surgical therapy. I. Report of the protocol and patient population. Am J Cardiology 37:896–902, 1976

37. Unstable Angina Pectoris Study Group: Unstable angina pectoris: National Cooperative Study to compare surgical and medical therapy. II. In-hospital experience and initial follow-up results in patients with one, two, and three-vessel disease. Am J Cardiology 42:839–848, 1978
38. Unstable Angina Pectoris Study Group: Unstable angina pectoris: National Cooperative Study Group to compare surgical and medical therapy. III. Results in patients with S-T segment evaluation during pain. Am J Cardiology 45:819–824, 1980
39. European Coronary Surgery Study Group: Coronary–artery bypass surgery in stable angina pectoris: Survival at two years. The Lancet 889–893, 1979
40. European Coronary Surgery Study Group: Prospective randomized study of coronary artery bypass surgery in stable angina pectoris. Second interim report. The Lancet 491–495, 1980
41. Kloster FE, Kremkau EL, Ritzmann LW, Rahimtoola SA, Rosch J, Kanarek PH: Coronary bypass surgery for stable angina. A prospective randomized study. 300:149–157, 1979
42. Barratt-Boyes BG: Cardiothoracic surgery in the antipodes. Including a report of a randomized trial of medical and surgical treatment of asymptomatic patients with severe coronary artery disease; a long-term follow-up of both fresh and antiobiotic-treated homograft valves; and some observations on glutaraldehyde-preserved valve tissue. J Thorac Cardiovasc Surg 78:804–822, 1979
43. Norris RM, Agnew TM, Brandt PWT, Graham KJ, Hill DG, Kerr AR, Lowe JH, Roche AHG, Whitlock RML, Barratt-Boyes BG: Coronary surgery after recurrent myocardial infarction: Progress of a trial comparing surgical with nonsurgical management for asymptomatic patients with advanced coronary disease. Circulation 63:785–792, 1981
44. Alderman EL, Fisher LD, Maynard C, Mock MB, Ringqvist I, Bourassa MG, Kaiser GC, Gillespie MJ: Determinants of coronary surgery in a consecutive patient series from geographically dispersed medical centers. The Coronary Artery Surgery Study. Circulation 66:Suppl. I, I-6–I-15, 1982
45. Kennedy JW, Kaiser GC, Fisher LD, Maynard C, Fritz JK, Meyers W, Mudd JG, Ryan TH, Coggin J: Multivariate discriminant analysis of the clinical and angiographic predictors of operative mortality from the collaborative study in coronary artery surgery (CASS). J Thoracic Cardiovasc Surg 80:876–887, 1980
46. Coronary Artery Bypass Surgery. National Institutes of Health Consensus Development Conference Summary. Volume 3, Number 8, 1981.
47. Meier PL: Quoted in Kolata GB: Coronary treatment assessed. Science 213:195, 1981
48. Preston TA: Invited Commentary. World J Surg 2:807–809, 1978

Overview of Recent Clinical and Methodological Advances from Clinical Trials of Cardiovascular Disease

William T. Friedewald and James A. Schoenberger

National Institutes of Health, Bethesda, Maryland

ABSTRACT: Concurrently controlled, randomized clinical trials play an important role in the validation of treatment and control measures of cardiovascular disease. As such trials have been carried to conclusion, a great deal has been learned not only about the treatment of the disease but also about the methodology of clinical trials. Seven recently completed clinical trials, the Coronary Drug Project, the Aspirin Myocardial Infarction Study, the Anturane Reinfarction Trial, the Hypertension Detection and Follow-Up Program and three drug trials of betablockers, are reviewed from the perspective of both their clinical and methodological significance.

KEY WORDS: *myocardial infarction, clinical trials, clinical trials methodology, secondary prevention, coronary heart disease*

INTRODUCTION

Randomized clinical trials play an important role in the validation of treatment and control measures of cardiovascular disease. As such trials have been carried to conclusion, a great deal has been learned not only about the treatment of the disease but also about the methodology of clinical trials. From these and other similar trials have evolved certain principles for the conduct of clinical trials that have received wide acceptance among investigators and that have been incorporated with increasing frequency in the design and analysis of ongoing and recently completed trials. Other concepts are still in an evolutionary phase and are less well accepted on a routine basis.

We do not attempt here to discuss all the clinical trials that have been carried out in the cardiovascular areas. We limit the discussion to seven completed major trials that illustrate some of the methodological problems and the lessons learned. The following trials have addressed major issues in

---

Received June 14, 1982; accepted July 1, 1982.

Address reprint requests to: Dr. William T. Friedewald, National Institutes of Health, Federal Building, Room 216, Bethesda, MD 20205, USA.

Controlled Clinical Trials 3:259–270 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030259-12$02.75

the management of cardiovascular disease and their findings, whether negative or positive, have had important effects on medical practice:

1. The Coronary Drug Project (CDP).
2. The Aspirin Myocardial Infarction Study (AMIS).
3. The Anturane Reinfarction Trial (ART).
4. The Hypertension Detection and Follow-Up Program (HDFP).
5. Three recent trials of beta blocker drugs (timolol, metoprolol, and propranolol).

CORONARY DRUG PROJECT

The Coronary Drug Project (CDP) [1] was instituted to investigate whether mortality during an extended follow-up period could be reduced through the use of drugs that lower serum cholesterol. Survivors of a myocardial infarction were chosen because they have a high annual mortality rate and thus the event rate would be sufficiently high to allow a manageable sample size and a reasonable follow-up period. Lowering of cholesterol was achieved by use of five different drug regimens. This study was one of the largest clinical trials ever attempted, as well as being one of the earliest clinical trials in the prevention of coronary heart disease (CHD).

Over 8000 postinfarction men were randomized into five equal-sized treatment groups and into a placebo group two and a half times the size of any other group. One of the earliest problems encountered in this trial was the difficulty in recruitment, a problem that has plagued every other clinical trial since that time. Because participants were not entered into the trial in the immediate postinfarction period, it was eventually found feasible and essential to make a direct public appeal for participants, bypassing direct physician referral, which had proven to be inadequate. The method was successful and has been used widely in subsequent trials when appropriate.

As the trial progressed it was closely monitored by two review boards of experts, a Data Safety and Monitoring Committee (DSMC) and a Policy Board, the members of which had no personal involvement in the trial. These groups periodically reviewed the progress of this double-blind trial for the purpose of detecting any toxic effects of the drugs used as well as determining whether the study was successful and could be terminated prior to the scheduled end as identified in the protocol.

The DSMC usually spent one day semiannually carefully reviewing the unblinded data in great detail and then made recommendations to the Policy Board. This board was then to consider the data, the recommendations of the DSMC, as well as new scientific developments and other policy issues. They then made a recommendation to the National Heart, Lung, and Blood Institute with regard to continuation of the study and specifically with regard to continuation or discontinuation of any of the various drug treatments.

Although the issue of a well-defined protocol was not a particular issue in the CDP, in other trials it has been. As with any good scientific experiment, one needs to set down carefully in advance the important elements of the research. In multicenter clinical trials, perhaps even more so than in routine laboratory experimentation, this step is critical. To the extent that a protocol is open ended, allowing for many possible end points, methods of analyses, inclusion or exclusion of patients, and an unspecified period of follow-up, one obviously does not have a very tightly designed scientific experiment for which one can be appropriately criticized later.

The need for a clearly defined primary end point is obvious. However, it has been argued that in addition to this requirement, the protocol should have clearly defined early stopping statistical rules. A number of us have resisted the routine inclusion of such "stopping rules" into the protocols because we do not think they are feasible. The decision to stop a clinical trial such as the CDP should never be based solely on having reached a certain *p* value from a statistical test of the data. The decision must be based on a number of other issues, e.g., toxicity and consistency. It is the role of the DSMC and Policy Board to weigh *all* of the relevant issues and come to an informed judgment. Still one of the more difficult areas that we have dealt with over the years in making these decisions has been the statistical problems created by multiple looks and multiple comparisons of the accumulating data and how to control for error under those circumstances. This problem led to a great deal of discussion within the CDP and several possible solutions were developed based on new statistical research. This research has obviously continued outside the CDP and new monitoring techniques have evolved, although no standard, uniformly acceptable procedure has been developed.

In the case of the CDP, three of the drug regimens (dextrothyroxine, low-dose estrogen, and high-dose estrogen) were discovered to be causing potentially more harm than benefit and led to early termination of these regimens [2–4]. None of these drug treatments was continued to the point of statistically significant evidence of harm since that was clearly not the goal of the study. Although these drugs had never been in wide clinical use for the management of CHD, the findings of the CDP were presumably of clinical importance in discouraging future use of these drugs in this context.

Of great utility were the lessons learned about the natural history of CHD [5] in the large placebo group of over 2700 patients. The characteristics that had been identified as predictive of the risk of initial development of CHD in many prospective studies were found to be relatively less important after myocardial infarction. Those characteristics that did seem to be associated with subsequent reinfarction and death were related rather to the amount of myocardial damage sustained in the prior attack, the amount of remaining ischemic myocardium, and the presence of ventricular irritability. These findings have supported treatment strategies in managing postinfarct patients aimed toward reduction of infarct size and prevention of cardiac dysrhythmia. The control of the traditional risk factors should clearly not be abandoned, but there is less expectation regarding the reduction in mortality that can be achieved by this means.

Finally, the CDP demonstrated that the two drug regimens, clofibrate and nicotinic acid [6], which were continued for the full five years of follow-up, were moderately successful in lowering serum cholesterol, but they did not have any impact on total mortality or cardiovascular incidence (nonfatal myocardial infarction and death due to infarction). These results for clofibrate, plus the findings of certain undesirable side effects, e.g., gallstones [7], probably added to a gradual reduction of the use of clofibrate in clinical practice.

This decline in use was almost certainly hastened by confirmatory reports of toxicity, this time mortality, from the primary prevention trial of clofibrate in Europe [8].

ASPIRIN MYOCARDIAL INFARCTION STUDY

Platelet aggregation plays an important role in thrombosis as well as in atherosclerosis. Theoretically, inhibition of platelet aggregation could retard intraarterial thrombosis, a presumed factor in myocardial infarction, and perhaps also retard atherogenesis. In addition, inhibition of thromboxane A2 by antiplatelet drugs in low dose could be of potential clinical value since thromboxane is a potent vasoconstrictor that could play a role in coronary vasospasm. Although some of this information was developed after the initiation of AMIS, the general concept did help form the theoretical basis in support of AMIS and similar trials of aspirin and other antiplatelet drugs. Later the information on endothelial prostacyclin, a potent platelet antiaggregator and vasodilator also inhibited by these drugs, became available and tended to confuse the theoretical basis for these trials. However, a great deal of clinical trial work had already been done in the late 1960s and early 1970s, forming a sound clinical rationale, even without a firm theoretical basis for undertaking these investigations.

The largest of these trials, AMIS [9], randomized over 4500 postinfarct patients into a placebo group or a treatment group receiving 1.0 g of aspirin daily. Both groups were followed for an average of about three years, with some individuals randomized into the trial shortly after their infarct, others up to 5 years later. Recruitment for this trial was successful in that the time allowed for obtaining the desired sample size was adequate and the goal was even exceeded by a few hundred participants. The recruitment techniques we learned over time in the CDP were put in effect immediately with AMIS and were remarkably successful.

In contrast to the other published major clinical trials on the use of antiplatelet agents in the postinfarction period, the results in AMIS clearly showed no evidence of benefit [10]. The other trials reported usually nonstatistically significant positive trends. In AMIS there was no benefit at all for prevention of the primary end point, death, and only a small nonsignificant difference for nonfatal myocardial infarction in favor of the aspirin group.

Because the other trials had shown favorable trends for antiplatelet therapy [11], the issue of the clinical usefulness of these drugs remains clouded. Adding to this is the problem of inhibition of prostacyclin by these agents. There is uncertainty whether there is a dose that will inhibit thromboxane and not inhibit prostacyclin. Since this dose relationship may vary remarkably by individual and for different vascular beds and whether the vessel wall is atherosclerotic or not, a great deal of laboratory and clinical research must be completed before future clinical trials are warranted with dosages lower than those already tested. Certainly the results of the completed clinical trials of aspirin do not support a dose-response difference within the dosage range tested (350–1500 mg/day).

Other lessons learned from AMIS dealt with assessment of adherence to the assigned drug regimen and the desirability of using multiple methods of ascertainment of adherence. The use of aspirin in a dose of 500 mg twice daily proved not to be innocuous, for there were significantly more side effects relative to placebo, especially involving the gastrointestinal system even on this relatively low dose of aspirin. An observed modest rise in BUN was not clinically important but worrisome in terms of possible danger of extended use of aspirin. A related issue is that adherence to the prescribed dosage is quite good in the clinical trial setting and remarkably better than one finds in a routine therapeutic mode. The compliance literature indicates that if a drug is prescribed as treatment for a chronic disease, the adherence rate is about 50%, whereas in the clinical trials we routinely obtain somewhere between 70% and 90% adherence. There must be lessons that can be taken from clinical trials and transferred to the therapeutic setting. Our own judgment is that a good portion of the clinical trial success is due to the high degree of attention given to adherence with frequent patient reminders, discussion of specific problems in taking drugs, and constructive attempts to work through the problems with the patients.

Other side issues investigated in AMIS were the quality of the informed consent process and the amount of information that was retained by the patient; the effectiveness of the double-blind design; the differences between the self-reported and laboratory-measured adherence with that revealed in a private open interview setting with a neutral person; and finally, the effectiveness with which information about the study was disseminated to the practicing physician and the impact, if any, that this trial had on the practice of medicine. The results of these methodological investigations have either been published or are in process.

In summary, AMIS was a large clinical trial that gave no support to the popular hypothesis that a small daily dose of aspirin will prevent recurrent CHD. How widely the public subscribes to the concept is not known and whether habits have changed as a result of AMIS is also uncertain. In the recruitment of AMIS participants it appeared that a number of physicians also subscribed to the theory that aspirin was useful.

ANTURANE REINFARCTION TRIAL

In the Anturane Reinfarction Trial (ART) many methodological issues arose following a preliminary report of the trial [12] and continued to surface following the publishing of the final results [13].

The trial evaluated in a randomized, double-blind design the effect of sulfinpyrazone in 1629 postmyocardial infarction patients randomly assigned to treatment or placebo 25–35 days after infarction. The objective in the protocol was "to compare the effects of Anturane (sulfinpyrazone) versus placebo on the rates of cardiac mortality associated with recurrent myocardial infarction in patients with a recent acute infarct." Despite this stated objective the trial claimed benefit mainly from a reduction in sudden death during the first six months after infarction [13]. Although the original rationale for the trial was the antiplatelet action of sulfinpyrazone, the investigators postulated that their dramatic result on this new end point was the consequence of possible antiarrhythmic properties of the drugs.

ART raised more issues about methodology than any other clinical trial in recent years. It is not known how many physicians began to prescribe sulfinpyrazone after the initial report nor how many were dissuaded from its use following critical reviews of the methodology and data analysis in subsequent reports. The fact that approval of the use of sulfinpyrazone after infarction was not given by the FDA prevented advertising to the medical profession and probably inhibited widespread use.

One of the major issues raised by ART was that of withdrawing patients after randomization into the trial. A number of withdrawals were performed. The first was based on nonconformity of some patients with the protocol baseline inclusion/exclusion criteria, so-called "ineligibles." The following questions were posed by ART: should such ineligibles be withdrawn from the study and thus the analysis of the data? If so, by whom should the decision to exclude an individual patient be made? Finally, when should this decision be made? Because of the opportunity for bias, we feel that such patients should not be withdrawn and the end-point results for these "ineligible" patients should be reported in the final paper. Withdrawal several months after randomization, and after the patient had died in some cases, by individuals at the pharmaceutical house sponsor seems totally unacceptable. The mortality experience in the 71 patients withdrawn from ART was statistically different between the two treatment groups (10 of the 38 withdrawn sulfinpyrazone patients died compared to 4 of the 33 withdrawn placebo patients) [14].

Another methodological issue posed by ART was that of excluding events from the numerator and patients from the denominator when calculating rates based on the patient's adherence to the prescribed treatment regimen. Thus, events that occurred within seven days of starting initial therapy or of restarting therapy after a temporary discontinuation were not counted (the seven days was chosen because this was the reported period required for the purported therapeutic effect of the drug to occur). All events while not on any therapy were also omitted from analysis. Data published from the Coronary Drug Project dramatically exemplify the problems that can be created by such practices. The mortality experience in the adherers to placebo in that trial was statistically significantly better than for the nonadherers to the placebo. We therefore feel that withdrawals based on adherence should not be performed because of the strong potential for bias. At a minimum, such withdrawals destroy the original randomization and the insurance that procedure provides in the interpretation of the results. Including nonadhering patients in the analysis dilutes the power of the trial, but adequate power can be maintained by increasing the sample size to adjust for this nonadherence.

Another issue raised by the ART publications was that not all of the data were presented in such a manner that readers could reproduce their analyses or perform additional analyses.

Obviously no report can present sufficient data to meet all possible needs, but a clear presentation of numerators and denominators as well as *p* values for the major events, major subgroups, for all patients randomized and for several relevant subsets should be a minimum requirement for such important studies.

Perhaps the most critical problem raised by the ART and identified in the paper from the FDA [14] was the problem of unbiased classification of events. This process should clearly rest entirely with groups that do not know the data trends or patient assignment. In addition, to the extent possible, unambiguous, mutually exhaustive and nonoverlapping definitions of events should be developed. On several counts the ART did not meet these goals.

Finally, the problem of drawing a firm conclusion as opposed to forming a new hypothesis when dramatically unexpected results develop in a study was clearly focused by the ART. The trial was originally designed to look at "cardiac mortality associated with recurrent myocardial infarction" with a minimum of 1 year follow-up and maximum of 2 years for each patient. In fact, the major finding of the study was a reported reduction in sudden death during the first 6 months of follow-up. Neither parameter had even been considered in advance of the trial. Obviously such unexpected results should not be ignored, simply because they were not anticipated in the design. However, instead of using this observed outcome to develop a new hypothesis to be confirmed potentially by other studies, the ART investigators claimed that they had clear evidence of a real effect.

HYPERTENSION DETECTION AND FOLLOW-UP PROGRAM

The Hypertension Detection and Follow-Up Program (HDFP) was mounted to address a number of important unanswered questions regarding the treatment of high blood pressure [15]. The Veterans Administration Studies [16,17] had shown the value of pharmacologic treatment of diastolic hypertension above 105 mm Hg measured in hospitalized patients, but left unanswered the value of treatment for persons with lower diastolic blood pressures, especially 90–104 mm Hg, and the demonstration of the value of treatment in ambulatory patients in a community setting, in younger patients, and in women.

The HDFP was a randomized, nonblinded clinical trial involving nearly 11,000 participants with diastolic blood pressure > 90 mm Hg after the second of two screening visits. These participants were recruited in such a manner as to maximize the opportunity for generalization of the results. Most centers in fact used home visits in an attempt to enter every person with elevated blood pressure within predefined census tracts. In addition there were almost no exclusions in the protocol. Patients with preexisting history of myocardial infarction, stroke, and the like were entered, as well as those with inadequately treated (according to the HDFP protocol) hypertension. Half of the patients were randomly referred to the usual health care sources in the community and had only minimal study follow-up in subsequent years in the form of an annual visit. The other half were seen frequently and intensively in special centers where aggressive antihypertensive drug treatment was administered according to a carefully defined stepped protocol. The rationale for using a nonplacebo control group was the perceived ethical dilemma created by the results of the VA Studies of placing hypertensives with diastolic blood pressure as high as 129 mm Hg on placebo treatment for the extended period of follow-up of five or more years. The compromise meant that the "control" group could also be on treatment although presumably much less vigorous and less frequent based on existing data from community health care resources and on the absence of a clear therapeutic mandate for vigorous treatment of the patient with hypertension. In actual fact, as the trial progressed, simultaneous programs aimed at increasing public and professional awareness of the importance of controlling hypertension led to ever more frequent treatment in the group referred to community care. Despite this handicap, blood pressure was more effectively lowered in the special centers and a significant difference between mean blood pressures for the two groups was maintained throughout the study.

The results revealed that more aggressive treatment of elevated blood pressure reduced mortality to a significant degree overall (17%) but to even a greater degree (23%) in the so-called mild hypertensive group (DBP 90–104) [18]. These results almost certainly understate the value of treatment since greater reductions could have been expected if there had been a placebo control group. The clinical implication of this important trial seems to be a consensus that persistent elevation of diastolic blood pressure above 90 mm Hg warrants treatment. This standard of clinical practice is now being introduced to the medical profession and can be expected to have a profound impact on medical practice in the future.

Some of the important treatment lessons of the HDFP were

1. the value and effectiveness of a carefully developed stepped care protocol;
2. the importance of the establishment of a goal blood pressure with continued aggressive therapy until that pressure is reached and maintained; and
3. the ability of a staff approach to the patient with nurse practitioners and other professionals implementing the stepped care protocol under the overall supervision of a physician to be so effective in lowering blood pressure.

Because the trial was nonblind and the treatment group was seen on a much more intensive and frequent basis at the specialized research centers, one of the concerns raised by critics has been whether the observed effect was due to the antihypertensive treatment of the patients or rather due to a higher level of clinical supervision and general care. The critics cite the observed reduction in noncardiovascular mortality in support of their criticism. The investigators have noted that high blood pressure is a risk factor for such noncardiovascular mortality. They have argued that HDFP showed striking results with the end points of stroke incidence and reversal and prevention of left ventricular hypertrophy that are highly unlikely to have occurred with nonspecific care alone. Therefore, they maintain, and we agree, that it is reasonable to attribute a significant part of the end-point results of HDFP directly to the lowering of blood pressure.

Perhaps the most important finding of the HDFP was the demonstration that mortality due to CHD was reduced by effective antihypertensive treatment. Previous VA and other studies had failed to demonstrate this, leaving some uncertainty about the risk factor concept of coronary atherogenesis. The increasingly effective identification and treatment of hypertensive individuals in the United States as a result of the National High Blood Pressure Education Program may explain, in part, the downward trend in cardiovascular mortality that has been observed in the past 10 years. Although stroke mortality has been declining for over 40 years, the rate of decline has been accelerated in the past 10 years.

In summary, the HDFP was a clinical trial with positive results and broad implications for the primary prevention of CHD and stroke. Vigorous dissemination of the findings should lead to a continual downward trend of cardiovascular mortality. In this sense, it should be viewed as a landmark study.

BETA-BLOCKER TRIALS

No other area of clinical trials in cardiovascular disease is of as much current interest as that of the use of beta blockers in the immediate postinfarction period based on the recent publication of the results of three large trials studying these drugs. All three trials were broadly similar in design, but used different beta blockers—timolol [19], metoprolol [20] and propranolol [21]. Patients were recruited at the time of hospitalization for their acute infarct and then followed up on medication an average of 3 months (metoprolol), 17 months (timolol), or 25 months (propranolol).

Together these trials revealed that the regular oral use of these beta blockers after acute myocardial infarction resulted in a significant reduction in total mortality, averaging about 35% at one year with evidence of continued effect as long as three years after infarct. Since the three different beta blockers were found to be equally useful, it seems reasonable that the observed effect is a class action due to beta blockade, perhaps blockade of the B1 receptors, since all the drugs share this property, rather than to separate unique properties of each drug.

Left unanswered by these trials is the value of therapy initiated after the immediate postinfarct period as well as the duration of benefit in long-term treatment.

Few methodological issues have been raised with regard to these recently reported studies. Unfortunately, the timolol study, despite highly significant results when all randomized participants were included in the analysis, concentrated the major portion of the paper to analysis based on the adherence to treatment principle. We hope repeated discussion of the problems with this form of analysis will lead to a reversal of the emphasis—with only a small amount or none of the paper devoted to analysis by adherence.

Time has not allowed a discussion of clinical trials of coronary bypass surgery. Yet one important lesson does emerge from the controversy over the role of this procedure in the treatment of coronary artery disease. The time to initiate a well-designed clinical trial is before medical practitioners and the public have prejudged the outcome and have made irrevocable therapeutic judgments. Such decisions may preclude a potentially important clinical trial, if for no other reason than that a randomization cannot be accomplished and therapeutic crossovers cannot be prevented.

In summary, the usefulness of clinical trials in cardiovascular disease has been demonstrated both by those with both neutral (CDP, AMIS) and positive (HDFP, beta blockers) results. Despite their complexity and cost, clinical trials remain the scientific cornerstone of decision making in prevention and treatment of cardiovascular disease.

REFERENCES

1. The Coronary Drug Project Research Group: The Coronary Drug Project: Design, methods, and baseline results (AHA Monograph No. 38). Circulation 47 (Suppl 1):11–150, 1973
2. The Coronary Drug Project Research Group: The Coronary Drug Project: Initial findings leading to modifications of its research protocol. J Am Med Assoc 214:1303–1313, 1970
3. The Coronary Drug Project Research Group: The Coronary Drug Project: Findings leading to further modifications of its protocol with respect to dextrothyroxine. J Am Med Assoc 220:996–1008, 1972
4. The Coronary Drug Project Research Group: The Coronary Drug Project: Findings leading to discontinuation of the 2.5 mg/day estrogen group. J Am Med Assoc 226:652–657, 1973
5. The Coronary Drug Project Research Group: Factors Influencing Long-term Prognosis After Recovery from Myocardial Infarction—Three Year Findings of the Coronary Drug Project. J Chron Dis 27:267–285, 1974
6. The Coronary Drug Project Research Group: Clofibrate and niacin in coronary heart disease. J Am Med Assoc 231:360–381, 1975
7. The Coronary Drug Project Research Group: Gallbladder disease as a side effect of drugs influencing lipid metabolism. New Engl J Med 296:1185–1190, 1977
8. Report from the Committee of Principal Investigators: A cooperative trial in the primary prevention of ischaemic heart disease using clofibrate. Brit Ht J 40:1069–1118, 1978
9. Aspirin Myocardial Infarction Study Research Group: Prostaglandins in hematology. In: *A Study of Aspirin in Patients with Myocardial Infarction*, Spectrum, 1977, pp 361–370.
10. The Aspirin Myocardial Infarction Study Research Group: A randomized controlled trial of aspirin in persons recovered from myocardial infarction. J Am Med Assoc 243:661, 1980
11. Braunwald E, Friedewald WT, and Furberg CD (Eds): *Proceedings of the Workshop on Platelet-Active Drugs in the Secondary Prevention of Cardiovascular Events.* AHA Monograph 74, 1980
12. The Anturane Reinfarction Trial Research Group: Sulfinpyrazone in the prevention of cardiac death after myocardial infarction. The anturane reinfarction trial. New Engl J Med 298:289–295, 1978
13. The Anturane Reinfarction Trial Research Group: Sulfinpyrazone in the prevention of sudden death after myocardial infarction. New Engl J Med 302:250–256, 1980
14. Temple R, Pledger GW: The FDA's critique of the anturane reinfarction trial. New Engl J Med 303:1488–1492, 1980
15. The Hypertension Detection and Follow-Up Program Cooperative Group: The Hypertension Detection and Follow-Up Program. Prev Med 5:207–215, 1976
16. Veterans Administration Cooperative Study Group on Hypertensive Agents: Effects of treatment on morbidity in hypertension: II. Results in patients with diastolic blood pressure averaging 90 through 114 mg Hg. J Am Med Assoc 213:1143–1152, 1979
17. Veterans Administration Cooperative Study Group on Hypertensive Agents: Effects of treatment on morbidity in hypertension: Results in patients with diastolic blood pressure averaging 115 through 129 mg Hg. J Am Med Assoc 202:1028–1034, 1967
18. Hypertension Detection and Follow-Up Program Cooperative Group: Five-year findings of the Hypertension Detection and Follow-Up Program. 1. Reduction in mortality of persons with high blood pressure, including mild hypertension. J Am Med Assoc 242:2562–2577, 1979

19. Norwegian Multicenter Study Group: Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. New Engl J Med 304:801–807, 1981
20. Hjalmarson A, Herlitz J, Malek I, et al: Effect on mortality of metoprolol in acute myocardial infarction. A double-blind randomized trial. Lancet 2:823–827, 1981
21. Beta-Blocker Heart Attack Trial Research Group: A randomized trial of propranolol in patients with acute myocardial infarction. 1. Mortality results. J Am Med Assoc 247:1707–1714, 1982

DISCUSSION FOLLOWING PRESENTATION

In response to questions concerning stratification in the CDP, Dr. Klimt said that stratification had been on clinics and on the severity of the infarct. This minimal stratification was used because of the expectation that there would be no more than 100 patients in the average size clinic and there was the danger of ending up with cells of an inadequate size with further stratification. This proved to be wise because it was learned in the course of the trial that one factor not stratified for—the New York Heart Association's classification of the disease—did not have the expected even distribution between class I and II, but was two-thirds class I and one-third class II.

Dr. Gehan pointed out that stratification on clinics usually was not done in cancer trials where the assumption is that the risk factors associated with the prognosis are sufficient. Dr. Meinert said there was a great variability in the mortality experience from clinic to clinic in the CDP with the highest mortality being two or three times the rate of the lowest and that a similar variability had been found in the UGDP.

Dr. Stamler said there had been a strong sentiment in favor of stopping the CDP trial when the cholesterol-lowering results of the drug regimens had been demonstrated. The drugs were to be declared "winners" and the results made known so that the public would benefit. It turned out of course, that the so-called "winners" had no positive effect at all in preventing secondary myocardial infarcts and there was later evidence of harmful side effects from the drugs. He further commented that there had been a lost opportunity when treatment was stopped from the group on dextrothyroxin and for the two groups on high and low dose estrogens. A special committee was formed to discuss possible new studies on these patients and to consider a number of possibilities. Beta blockers were considered but the committee decided the risks were too high and voted not to do a beta blocker study. Aspirin was another possibility and the study of aspirin was implemented. Dr. Friedewald added that there had been discussions when the trial was first designed of adding aspirin to the nicotinic acid regimen to prevent the itching and flushing that nicotinic acid caused. Ultimately, the view prevailed that even though it was "just aspirin" it was not known what its effect would be and it was not used.

There was a discussion of what should be concluded from the series of trials on aspirin therapy after myocardial infarct. Dr. Chalmers surmised there might be a 10% reduction in annual case fatality rate using aspirin and that this would be undetectable by the studies which had been done, including

AMIS. Considering the number of deaths from subsequent myocardial infarcts, this would be an extremely important result. Dr. Klimt pointed out that the AMIS study was not significantly different from the other positively tending studies. Thus, even though it was on the other side of the spectrum, it was not significantly different from the other aspirin studies. Dr. Schoenberger added there were unknown side effects to aspirin which weigh against its chronic use.

Overview of Randomized Clinical Trials of Lipid Intervention for Atherosclerotic Cardiovascular Disease

Henry Buchwald, Laurie Fitch, and Richard B. Moore

Department of Surgery, University of Minnesota Medical School, University of Minnesota Hospitals, Minneapolis, Minnesota

ABSTRACT: This paper is a review of the scientific and economic (when possible) impact of 15 randomized clinical trials of lipid intervention for atherosclerotic cardiovascular disease. No effort has been made to be all-inclusive and certain smaller and first generation studies have been omitted. No effort has been made to offer detailed analyses of protocols; we have limited discussion to summaries of the salient features of each study presented. After stating the data, we introduce certain reflections and raise some of the more obvious questions that emerge.

KEY WORDS: *randomized clinical trials, atherosclerotic cardiovascular disease, lipid intervention*

SCIENTIFIC SUMMARY

Oslo Diet Heart Study

One of the very first clinical trials of the effect of hypolipidemic therapy on coronary heart disease was conducted by Leren [1] in Norway between 1956 and 1963. This secondary (i.e., after overt manifestation of atherosclerotic coronary artery disease) prevention trial consisted of 412 men aged 30–64 years, who had sustained a first myocardial infarction 1–2 years previously. They were randomized to either a modified fat-controlled diet (206) or a normal unmodified diet (206) and followed for a period of five years. The primary end point was myocardial infarction and secondary end points were new cases of angina pectoris and sudden death. The mean cholesterol reduction achieved in the modified diet group was 17.6% versus a 3.7% reduction in the control group, giving a net 13.9% difference in plasma cholesterol. For all ages combined there was a significant reduction in the incidence of myocardial infarction ($p = 0.02$) and in newly developed angina pectoris ($p = 0.03$); however, the number of cases of sudden death was the same in

---

Received June 14, 1982; accepted July 1, 1982.

Address reprint requests to: Dr. Henry Buchwald, Department of Surgery, University of Minnesota Medical School, Box 290, University of Minnesota Hospitals, Minneapolis, MN 55455, USA.

Controlled Clinical Trials 3:271–283 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030271-13$02.75

each group. When dividing the subjects into those 60 years of age and older and those younger than 60 years of age, only the younger age group showed a significant reduction in myocardial infarction and new angina, and again there was no difference in the rate of sudden death. The investigators continued to follow the trial subjects for mortality (but not for adherence to diet or serum lipids levels) for several years after the completion of the trial and an 11-year follow-up report sustained the original observations [2].

Veterans Administration Unsaturated Fat Study

Dayton et al. [3] conducted a randomized double-blind combined primary (i.e., before overt manifestation of atherosclerotic cardiovascular disease) and secondary intervention study of the effect of a diet high in unsaturated fat between September 1959 and January 1968, using 846 volunteer males living in a Veterans Administration domicile. The subjects' ages ranged from 55 to 89 years, with an average age of 65.5 years. The primary end point of the study was coronary heart disease as exhibited by sudden death or acute myocardial infarction. Subject follow-up was maintained for up to 8 years following entry into the trial. The mean difference in serum cholesterol between the control and the diet group over the course of the trial was 12.7% but the difference in the primary end point of the study was not significant between the two groups. When these data are pooled with data from secondary end points (cerebral infarct, amputation due to ischemia, and so on), the modified diet group showed significantly ($p = 0.02$) fewer total events and fewer fatal atherosclerotic events. The investigators, however, concluded that their trial had not provided definitive answers to the questions of the effect of plasma cholesterol reduction on cardiovascular disease incidence or mortality. A later report dealt with the excess of deaths due to noncardiovascular causes in the modified diet group, especially the excess of deaths due to carcinoma, but the investigators did not feel that the excess noncardiovascular deaths in the modified diet treated group was due to the experimental diet high in unsaturated fat [4].

British Corn Oil Study

Corn oil (80 g/day) as a mode of hypolipidemic intervention therapy was used in one of the first randomized secondary trials of ischemic heart disease, reported by Rose, Thompson, and Williams [5] in 1965. A total of 80 subjects (sex not specified) were randomized to one of three groups: control ($n = 26$); olive oil supplement ($n = 26$); and corn oil supplement ($n = 28$). All subjects were less than 70 years of age, with the mean ages for the three groups being 58.8, 55.0, and 52.6 years, respectively. The length of follow-up was originally planned to be 3 years but due to the drop-out rate, the data were analyzed after only 2 years. The major end points of this trial were sudden death, fatal myocardial infarction, nonfatal myocardial infarction, angina, and "other significant cardiac pain." There was no change in the serum cholesterol levels of the subjects in the control group or in the olive oil group over the course of the trial; there was a significant ($p < 0.01$) reduction in the serum cholesterol in the corn oil group at 6 months, 1 year, and 1 and one-half years, but at 2 years this reduction was no longer statistically significant. Deaths occurred 25% more frequently in the corn oil group when compared to the control group. The percentage of subjects remaining in this trial that were free of major cardiac events (as a proportion of those subjects still in the trial at two years) was 75% for the control group, 57% for the olive oil group, and 52% for the corn oil group. The investigators subsequently did not recommend corn oil as a treatment for ischemic heart disease.

British Research Committee Soya-Bean Oil Study

In this study [6] conducted between 1960 and 1967, 199 males were allocated to a soy-bean-oil-supplemented low-fat diet group and 194 were allocated to a normal diet group; the subjects were followed for a period of 2–6$^3/_4$ years (average of 4 years), with half the men remaining in the trial for longer than 4 years. For this secondary prevention trial, the primary end point was cardiovascular relapse—evidenced by death from cardiovascular causes, nonfatal reinfarction, or angina. There was an immediate decline in serum cholesterol level in both the diet (−22%) and control (−6%) groups, which reached maximum reduction at 6 months after entry into the trial; thereafter, there was a steady increase in serum cholesterol in both groups, and at 5 years the diet group was only 12% below its starting level. However, over half of the subjects in the trial were on weight reduction diets. The investigators found no significant differences between the end points in the diet group and the control group and concluded that there was little support for dietary therapy in the treatment of patients who have suffered a myocardial infarction.

British Research Committee of London Diet Study

This controlled secondary clinical trial was conducted from 1957 to 1963 in three British hospitals [7]. A modified low-fat diet was the means chosen for intervention in 252 males under age 65 who were randomly allocated to either the control group (normal diet, $n = 123$) or the treatment group (low-fat diet, $n = 129$). The two major end points of the study were reinfarction and death from coronary heart disease; mean follow-up time was 3 years for both groups. At 6 months after entry into the trial, the diet-treated group had serum cholesterol levels significantly ($0.02 < p < 0.05$) lower than the control group. However, by 1 year and for the duration of the trial, this difference was no longer significant due to the continued decline of serum cholesterol level in the control group. At 3 years after entry into the trial, the reduction in serum cholesterol was 17% for the treated group and 9% for the control group. Both the control and the diet groups lost weight during the course of the trial. The investigators found no significant differences in the primary end points of the study between the two groups and therefore advised that the low fat diet had no place in the treatment of myocardial infarction.

Scottish Physicians Clofibrate Study

A secondary prevention trial using clofibrate (1.6–2.0 g/day) was conducted by a group of Scottish physicians between August 1964 and April 1970 [8]. Three hundred and fifty subjects (288 men, 62 women) received the drug and 367 (305 men, 62 women) received placebo. All subjects were between the ages of 40 and 69 years and had clinical evidence of coronary heart disease (recent myocardial infarction, or angina with positive ECG findings). End points were sudden death, fatal myocardial infarction, and nonfatal definite and nonfatal probable myocardial infarction. The follow-up period was up to 6 years, with an average length of 1.8 years. The serum cholesterol reduction achieved was 16% for both sexes in the last two years of the trial for the double-blind cohort (a subgroup of subjects receiving anticoagulants was not double blind); however, the placebo group's initial serum cholesterol was significantly ($p = 0.01$) higher than the clofibrate group's initial serum cholesterol. There was no significant overall reduction for the clofibrate-treated group in either sudden deaths or deaths from all causes in this trial. There was a significant reduction in mortality in subjects who had previously reported having angina pectoris but this effect was independent of the initial serum cholesterol level. Deaths due to noncardiac causes were excluded in the assessment of the results. The study concluded that clofibrate had a beneficial effect on mortality and possibly morbidity in patients who have angina pectoris but not in patients who had sustained a myocardial infarction without prior angina.

Newcastle on Tyne Clofibrate Study

This 5-year double-blind secondary prevention trial of clofibrate by an association of physicians in Britain was conducted between February 1964 and August 1969 [9]. A group of 244 subjects (192 men, 52 women) received 1.5–2.0 g/day of clofibrate, while a group of 253 subjects (208 men, 45 women) received placebo (comparable capsules of corn oil). All subjects were less than 65 years of age with an overall average age of 52 years. End points were cardiac death and first new nonfatal myocardial infarction; deaths due to noncardiac causes were excluded. The serum cholesterol reduction achieved in the clofibrate-treated group (as calculated from data in the report) was 11% in males and 15% in females after 6 months of intervention, and this trend continued throughout the trial. No figures are given for the placebo group. At the conclusion of the trial the findings were (1) fewer subjects in the clofibrate group died due to cardiac death ($p = 0.02$) and sudden death ($p = 0.03$), and (2) fewer of these clofibrate-treated subjects had non-fatal infarcts ($p = 0.055$). However, when subjects taking anticoagulants were removed from the analysis, there was no significant difference in sudden deaths but the difference in the rate of nonfatal infarction became more significant ($p = 0.03$). The risk of death from first new infarcts was not different between the clofibrate and placebo groups. More subjects taking clofibrate withdrew from the trial (44) than those taking the placebo (29), but there were no significant differences in the characteristics of the two withdrawal groups.

Minnesota Coronary Survey

A combination primary and secondary intervention trial was conducted by Frantz et al. [10] and Brewer, Ashman and Kuba [11] between October 1966 and June 1973 and consisted of 9449 institutionalized males and females between the ages of 21 and 102. The mode of intervention in this NIH-funded, investigator-initiated grant was a low-cholesterol, low-fat diet (4698 subjects) versus a normal institutional diet (4751 subjects). The follow-up period ranged from less than 1 year to a possible maximum of $4^1/_2$ years, with an average follow-up time of one year. The primary end point was death from all causes; other end points were cause-specific mortality and myocardial infarction. At the conclusion of the trial, the dietary group had achieved an average serum cholesterol reduction of 13.8%, while the control group sustained a 0.6% reduction. The trial found a statistically significant reduction in total deaths and cardiovascular events in men less than 50 years of age treated with modified diet, but these differences were not observed in the female diet-treated group. When subdividing into age groups, this difference in total deaths and cardiovascular events was still observed in the men but the differences were no longer statistically significant.

Coronary Drug Project

This large secondary prevention trial conducted between March 1966 and August 1974 was a national collaborative NIH-funded grant to investigate the effects of estrogen, dextrothyroxine, clofibrate, and niacin on coronary heart disease [12]. Estrogen and dextrothyroxine were discontinued before the completion of the trial due to adverse and at times lethal effects. The completed portion of the study consisted of males, aged 30–64 years, who were randomized to a 1.8-g/day clofibrate group ($n = 1103$), 3.0-g/day niacin group ($n = 1119$), or a placebo group ($n = 2789$). Average length of follow-up was 74 months (range 5–8 and one-half years) and the primary end point was total mortality. After correction for changes in serum cholesterol over time in the placebo group, the average reduction in serum cholesterol was 6.5% in the clofibrate group and 9.9% in the niacin group. The results of the trial showed no evidence of an effect of clofibrate on total mortality or cause specific mortality (which was contrary to the findings in the Newcastle [8] and Scottish [9] studies) and indeed there were excesses of thromboembolism, angina pectoris, gallstones, intermittent claudications, cardiac arrhythmias, and nonfatal cardiovascular events in the clofibrate group. The drug was therefore not recommended for use in patients with coronary heart disease. Niacin was found to be slightly beneficial in protecting against recurrent nonfatal myocardial infarction but there were no differences in total mortality or cause-specific mortality. There were excesses of arrhythmias, gastrointestinal problems, and abnormal blood chemistries with the use of this drug, and the investigators advised caution in the use of niacin for cholesterol reduction in patients having clinical coronary heart disease.

European Cooperative Clofibrate Study

This multicentered cooperative trial in the primary prevention of ischemic heart disease using 1.6 g/day of clofibrate was conducted between 1965 and 1976 [13]. Males, aged 30–59 (average age 45 years) were randomized into three groups: high-cholesterol clofibrate treated ($n = 5331$), high-cholesterol placebo treated ($n = 5296$), and low-cholesterol placebo treated ($n = 5118$). The average length of follow-up was 5.3 years. The primary end point was fatal ischemic heart disease and the expected reduction in serum cholesterol was estimated to be 15%. At the end of the trial the actual reduction in serum cholesterol achieved was 9%. The overall incidence of major ischemic heart disease was found to be significantly lower in the clofibrate treated group, but this difference was confined to the incidence of nonfatal myocardial infarction. Total mortality was significantly higher in the clofibrate-treated group on comparison to the placebo group (with the excess deaths due primarily to diseases of the liver, biliary tract, and intestine) although the age-standardized mortality rates did not differ significantly among the three groups. The investigators concluded that they could not recommend clofibrate for the primary prevention of coronary heart disease.

Upjohn Colestipol Hydrochloride Study

A combined primary and secondary single-blind randomized trial of colestipol hydrochloride (Colestid®) in males and females greater than 18 years of age and with serum cholesterol levels greater than 250 mg/100 ml was performed between 1969 and 1975 [14]. Fifteen grams per day of colestipol was given to the treatment group (548 men, 601 women) and 6 g per day of inert placebo was given to the control group (546 men, 583 women); the subjects were followed for 1–3 years. Major end points were total and cause-specific mortality. In subjects treated with colestipol for 3 years, the average reduction in serum cholesterol was 14%; in the control subjects over the same time period, the reduction was 5%. Coronary heart disease mortality in men was significantly ($p \leqslant .02$) less in the colestipol-treated group but total mortality was not significantly different ($p = 0.12$). There were no significant differences for women. Significant ($p = 0.05$) side effects reported were constipation, nausea, and exacerbation of hemorrhoids.

Swedish Combined Clofibrate and Nicotinic Acid Study

In a randomized clinical trial begun in December 1972 of 558 male and female survivors of a myocardial infarction, 229 were treated with a combination of clofibrate (2 g/day) and nicotinic acid (3 g/day); the remaining 229 subjects were allocated to the control group [15]. Mean age of the subjects was 59 years and all subjects were less than 70 years of age at entry into the trial. The major end points of the study were total mortality, ischemic heart disease mortality, and nonfatal myocardial infarction. Follow-up of the subjects continues at the present time but preliminary analysis has shown serum cholesterol to be lowered 15%–20% in the treated group at 6–36 months after entry into the trial. There was no difference found between the two groups for either total mortality or ischemic heart disease mortality but the number of nonfatal infarcts was reduced by 50% in the treated group. The authors suggest that the difference shown in the rate of nonfatal reinfarction (which is contrary to the findings of the Coronary Drug Project [12]) is due to the degree of serum lipid lowering obtained in this trial.

Coronary Primary Prevention Trial—Lipid Research Clinics (LRC)

In 1971, the National Heart, Lung, and Blood Institute initiated a primary prevention trial to test the lipid hypothesis [16]. Contracts were awarded to 12 clinical centers, each with its own lipid lab and ECG facility, and to 6 support centers. This is the most expensive trial ever attempted in the history of intervention trials. Type II hyperlipidemic males (cholesterol greater than 265 mg/100 ml and triglycerides less than 300 mg/100 ml) between the ages of 35 and 59 were randomly allocated to either a cholesterol-lowering diet and placebo group ($n = 1905$) or to a cholesterol-lowering diet and cholestyramine (24 g/day) group ($n = 1905$). Each participant is to have a minimum of 7 years of follow-up study. Primary end points are coronary heart disease mortality and nonfatal myocardial infarction, with numerous other secondary end points to be recorded. The published expected reduction in serum cholesterol was 4% in the control (diet only) group and 28% in the treatment (diet plus cholestyramine) group, for a net anticipated serum cholesterol reduction of 24%. The trial is currently in progress and no results have been published. However, it is our understanding that the LRC trial is currently achieving approximately one-half of their goal, i.e., 12%–14% net serum cholesterol reduction. Scheduled completion date of the trial is 1983.

Multiple Risk Factor Intervention Trial (MRFIT)

This National Heart, Lung, and Blood Institute initiated clinical trial began in 1972 and is scheduled for completion in 1982. Contract funding for 22 clinical centers and 4 support facilities make this project one of the largest in terms of number of subjects, manpower, and expense in the history of clinical trials of the lipid hypothesis. Males, aged 35–57 years with no evidence of heart disease but considered at high risk (diastolic blood pressure higher than 90 mm Hg, and/or serum cholesterol greater than 260 mg/100 ml, and/or smokers) according to the Framingham study risk predictions, have been randomized to an intervention group ($n = 6428$) or a nonintervention group ($n = 6433$). Intervention is to be accomplished by dietary modification, control of high blood pressure, and cessation of cigarette smoking. Follow-up is to be conducted for a minimum of 6 years. The expected reduction in serum cholesterol was 10% but the most recently published result of achieved serum cholesterol reduction is 7.4% at 4 years after entry into the trial. The primary end point is death due to coronary heart disease; other major end points are nonfatal myocardial infarction, death from all cardiovascular events, and death from all causes. Data concerning the end points of the trial have not yet been published.

Program on the Surgical Control of the Hyperlipidemias (POSCH)

POSCH is an investigator-initiated randomized secondary prevention clinical trial funded by a grant from the National Heart, Lung, and Blood Institute [18]. Initial work (pilot phase) began in 1973. There are a total of four clinical centers, with one centralized lipid lab. All support facilities are located at the University of Minnesota, the originating center. One thousand subjects are planned for the trial; the trial is still in the recruitment phase. Subjects are male and female volunteers, between the ages of 30 and 64 years, who have sustained one (and only one) documented myocardial infarction within the preceding five years, and who have serum cholesterol levels of 220 mg/100 ml or greater or low density lipoprotein cholesterol levels of at least 140 mg/100 ml if the serum cholesterol level is between 200 and 219 mg/100 ml. Subjects are not accepted into the trial with secondary hyperlipidemia; individuals with hypertension, diabetes, obesity, and other major diseases are excluded. The mode of intervention is partial ileal bypass, a surgical procedure designed to lower circulating blood cholesterol levels by an average of 40% and provide essentially obligatory intervention protocol adherence, i.e., minimal dropout from treatment. Up to 500 subjects will be randomized to the control group and up to 500 subjects will be randomized to the surgical intervention group; all subjects are to be followed a minimum of 5 years. The primary end point of the trial is death from all causes; secondary end points are fatal and nonfatal myocardial infarction, serial coronary arteriography changes, doppler assessment of peripheral occlusive disease, and new ECG abnormalities. Although the trial is still in progress, the reported serum cholesterol reduction in the surgical group at three years is 31% [19,20]. Significantly, the LDL-cholesterol fraction is reduced 44% at three years ($p < 0.05$) and there was a net 10% difference in the HDL-cholesterol between the surgery (−1.1%) and the control (−11.1%) groups. This trial is scheduled for completion in 1988.

COST SUMMARY

For the early clinical trials in both the United States and Europe, and the more recent trials in Europe, no financial data are available. We have been able to obtain financial information on the following trials: Minnesota Coronary Survey (MCS), Coronary Drug Project (CDP), Coronary Primary Prevention Trial (LRC), Multiple Risk Factor Intervention Trial (MRFIT), and the Program on the Surgical Control of the Hyperlipidemias (POSCH). Two of the above trials have been completed (MCS and CDP); the remaining are in progress.

The Minnesota Coronary Survey was funded by the National Heart, Lung, and Blood Institute (NHLBI) as an investigator-initiated grant and was conducted between 1967 and 1973 at a total cost of $6,000,000. The Coronary Drug Project, which was also a grant of the NHLBI, was active between 1966 and 1974 and the total cost was $41,500,000. LRC and MRFIT are NHLBI contracts with obligations through fiscal year 1980 of $99,859,176 for MRFIT and $126,144,589 for LRC. Total funds allocated to the present date for LRC are $142,864,000 through June of 1982; estimated total cost for MRFIT to the planned completion of this trial in 1982 are $115,769,176. For the POSCH trial the allocated funding through June of 1985 is $25,899,337; the estimated costs for this trial to the scheduled date of completion (June 1988) are $30,399,337.

Using the figures given above and the number of recruited and planned subjects for these five studies, the cost per subject in the trial is $635 for the MCS, $8,282 for the CDP, $37,270 for the LRC, $9,019 for MRFIT, and $30,000 for POSCH. In order to bring the dollar figures for the MCS and the CDP into line with the more recent LRC, MRFIT, and POSCH studies, we can (roughly) double the costs of the earlier studies. This will now allow a ranking by cost per subject in the trial for these five studies: MCS ($1,270), MRFIT ($9,019), CDP ($16,564), POSCH ($30,000), and LRC ($37,270). The MCS, CDP, LRC, and MRFIT trials employ diet and/or drug intervention; the POSCH study provides for surgical intervention in one-half of the subjects and three sequential coronary arteriograms for all of the subjects.

Table 1 summarizes the pertinent data of this review.

REFLECTIONS AND QUESTIONS—1982

1. All completed randomized clinical trials of lipid intervention for atherosclerotic cardiovascular disease have shown no convincing evidence for disease retardation, arrest, or reversal associated with plasma cholesterol reduction, although in no trial has cholesterol reduction been marked and in many it has been miniscule.
2. All trials, except for POSCH, have employed either or both diet and drug therapy for intervention.
3. The net mean calculable cholesterol reduction achieved by the diet and drug trials has been nine percent without the LRC and the Swedish Combined Clofibrate and Nicotinic Acid studies; if the presumptive reductions of these two studies are included, this average is 10%.
4. The net maximum cholesterol reduction achieved by the completed diet and drug trials is as high as 14%; the net maximum cholesterol reduction currently reported by the ongoing diet and drug trials is that of the Swedish Combined Clofibrate and Nicotinic Acid Trial and is in the 15%–20% range.
5. The only trial utilizing surgical intervention, the POSCH, is reporting a sustained 31% total plasma cholesterol reduction at three years, with a 44% reduction in the LDL-cholesterol and a 10% net higher HDL-cholesterol in the surgery group. This study will exhibit the maximal intervention modality adherence for all of the lipid trials.
6. A positive statement for intervention in a primary trial would not allow for a logical transfer of conclusions to secondary intervention; however, a positive statement for intervention in a secondary trial would allow an eminently logical extension of conclusions strongly supportive of primary intervention as well.
7. Lipid intervention trials performed under NIH-initiated contracts (in distinction to investigator-initiated grants) have been the most expensive to date.

Table 1 Summary of Randomized Clinical Trials of Plasma Cholesterol Reduction

Trial	Type of trial[a]	Dates	Number of subjects	Sex	Age range	Mode of intervention[b]	Duration of follow-up	Cholesterol[c] change	End points[d]	Results[e]
Oslo Diet-Heart Study	S	1956–63	412	M	30–64	D	5 yr	−17.6% (T) − 3.7% (C)	(1) MI (2) AP,SD	Subj < 60 yr: ↓MI & AP; SD (NS) Subj > 60 yr: NS
VA Diet Study	P&S	1959–68	896	M	55–89	D	≤ 8 yr	−12.7%	(1) SD or MI (2) CSM, CVA, Others	NS
British Corn Oil Study	S	→1965	80	—	<70	D	2 yr	−10%	SD, FMI, MI, AP	T: ↑SD, ↑MI
British Soya Oil Study	S	1960–67	393	M	<60	D	2–6.75 yr	−12%	MI, CVM, ECG	NS
British Diet Study	S	1957–63	252	M	<65	D	3 yr	−16.9% (T) − 9.4% (C)	MI, CM	NS
Scottish Clofibrate Study	S	1964–70	717	M&F	40–69	C	≤ 6 yr	−16%	SD, FMI, MI	Subj with AP: ↓SD, ↓FMI, ↓MI Subj without AP: NS
Newcastle Clofibrate Study	S	1964–69	497	M&F	<65	C	5 yr	−11% M −15% F	CM, MI	↓CM & ↓MI
Minnesota Coronary Survey	P&S	1967–73	9,449	M&F	21–102	D	< 1–4.5 yr	−13.2%	TM, CSM	Men < 50 yr: ↓TM & ↓CSM Men > 50 yr: NS

Coronary Drug Project	S	1966–74	5,011	M	30–64	C,N	5–8.5 yr	−6.5% C −9.9% N	(1) TM (2) CSM, MI, AP	All women: NS Men: NS for either drug
European Clofibrate Trial	P	1965–76	15,745	M	30–59	C	$\bar{x}$ = 5.3 yr	(−15% EX) −9%	CM, MI, TM, AP, ECG	↑ TM
Swedish Clofibrate/ Nicotinic Acid Study	S	1972–	558	M	<70	C + N	3+ yr	−15–20%	TM, CM, MI	TM & CM = NS ↓ MI
Upjohn Cholestipol Study	P&S	1969–75	2,278	M&F	>18	CP	1–3 yr	−9%	TM, CSM	Males: ↓ CSM Females: NS
CPPT (LRC)	P	1971–(1983)	3,810	M	35–59	D + CS	≥ 7 yr	(−24% EX)	(1) CM, MI (2) ECG, Others	(Trial in Progress)
MRFIT	P	1972–(1982)	12,861	M	35–57	D + BP + SM	≥ 6 yr	(−10% EX) −7.4% at 4 yr	CM, MI, TM, Others	(Trial in Progress)
POSCH	S	1973–(1988)	1,000	M&F	30–64	D + PIB	≥ 5 yr	(−40% EX) −31% at 3 yr	(1) TM (2) FMI, MI, ANGIO, AP, ECG	(Trial in Progress)

---

[a] P = primary, S = secondary, P&S = Combined primary and secondary.

[b] D = diet, C = clofibrate, CP = colestepol, N = niacin, CS = cholestyramine, BP = blood pressure control, SM = smoking control, PIB = partial ileal bypass surgery.

[c] T = treatment group, C = control group, EX = expected changes, C = clofibrate, N = niacin, M = males, F = females.

[d] End Points: (1) = primary, (2) = secondary, MI = acute myocardial infarction, FMI = fatal myocardial infarction, SD = sudden death, CM = cardiac mortality, TM = total mortality, CSM = cause-specific mortality, AP = angina pectoris, ECG = ECG changes, ANGIO = coronary arteriographic changes, CVA = cerebrovascular accident.

[e] Results: NS = no significant difference between T & C groups.

8. Should a government tax-supported agency—the NIH—perform drug studies? Should such studies, if attempted, insist on a strong financial contribution from industry?
9. Is rigid, recurrent peer review and investigator-initiated grants a proper mechanism for the funding of intervention trials or should such undertakings be given a protected, privileged status under contract and overall NIH control?
10. What is the best mechanism for goal-effective, patient-effective, science-effective, and cost-effective funding of intervention trials?

REFERENCES

1. Leren P: The effect of plasma cholesterol lowering diet in male survivors of myocardial infarction. Acta Medica Scandinavica Suppl 466, 1966
2. Leren P: The Oslo Diet-Heart Study. Eleven year report. Circulation 42:935, 1970
3. Dayton S, Pearce ML, Hashimoto S, Dixon WJ, Tomiyasu U: A controlled clinical trial of a diet high in unsaturated fat. Circulation 25(I):1, 1969
4. Pearce ML, Dayton S: Incidence of cancer in men on a diet high in polyunsaturated fat. Lancet 1:464, 1971
5. Rose GA, Thompson WB, Williams RT: Corn oil treatment of ischemic heart disease. Brit Med J 1:1531, 1965
6. British Research Committee, Morris, JN, Chairman: Controlled trial of soya bean oil in myocardial infarction. Lancet 2:693, 1968
7. British Research Committee, Ball KP, Chairman: Low-fat diet in myocardial infarction. Lancet 2:501, 1965
8. Research Committee of the Scottish Society of Physicians: Ischemic heart disease: a secondary prevention trial using clofibrate. Brit Med J 4:775, 1971
9. Physicians of Newcastle upon Tyne: Trial of clofibrate in the treatment of ischemic heart disease. Brit Med J 4:767, 1971
10. Frantz ID, Dawson EA, Kuba K, Brewer ER, Gatewood LC, Bartsch GE: The Minnesota Coronary Survey: Effect of diet on cardiovascular events and deaths. Circulation 52:(II)4, 1975
11. Brewer ER, Ashman PL, Kuba K: The Minnesota Coronary Survey: Composition of the diets, adherence and serum lipid response. Circulation 52(II):269, 1975
12. Coronary Drug Project Research Group: Clofibrate and niacin in coronary heart disease. J Am Med Assoc 321:360, 1975
13. Oliver MF, Heady JA, Morris JN, Cooper J: A cooperative trial in the primary prevention of ischemic heart disease using clofibrate. Brit Heart J 40:1069, 1978
14. Dorr AE, Gunderson K, Schneider JC Jr, Spencer TW, Martin WB: Colestipol hydrochloride in hypercholesterolemic patients-effect on serum cholesterol and mortality. J Chronic Dis 31:5, 1978
15. Carlson LA, Danielson M, Ekberg I, Klintemar B, Rosenhamer G: Reduction of myocardial reinfarction by the combined treatment of clofibrate and nicotinic acid. Atherosclerosis 28:81, 1977
16. Lipid Research Clinic Group: The Coronary Primary Prevention Trial: Design and Implementation. J Chronic Dis 32:609, 1979
17. Caggiula AW, Christakis G, Farrand M, Hulley SB, Johnson R, Lasser NL, Stamler J, Widdowson G: The Multiple Risk Factor Intervention Trial IV. Intervention on blood lipids. Prev Med 10:443, 1981
18. Buchwald H, Moore RB, Varco RL: Maximum lipid reduction by partial ileal bypass: A test of the lipid-atherosclerosis hypothesis. Lipids 12:53, 1977

19. Moore RB, Buchwald H, Varco RL, and the POSCH participants: The effect of partial ileal bypass on plasma lipoproteins. Circulation 62:469, 1980

20. Buchwald H, Moore RB, Rucker RD, Varco RL: Metabolic surgery: Management of morbid obesity and hyperlipidemia. Chapter In: *Surgery, Basic Principles, and Practice,* Hardy JD (Ed), in press.

A Potpourri of RCT Topics

Thomas C. Chalmers

Mt. Sinai School of Medicine, New York, New York

ABSTRACT: Five concepts were presented for discussion:

(1. *Randomization of the first patient.* At least six examples are known to the author, and there are undoubtedly others.

(2. *Factorially structured treatments.* The advantages of conducting two trials in the same patients when there is no suggestion of interaction, and documenting interaction if it exists must be weighed against the potential loss of power and the inconvenience of missing values.

(3. *RCT control groups as a source of unnatural history.* Such a small percentage of apparently suitable patients are actually randomized in most clinical trials that the control group can never be considered representative of that disease seen by practicing physicians.

(4. *Bias in random and nonrandom treatment assignments.* Differences in risk factors documented before randomization can be as much responsible for differences in outcome as the treatments under study.

(5. *Time lags between innovation, RCTs, and practice.* These have varied from a very few years (coronary bypass operations) to 154 years (treatment of scurvy).

DR. CHALMERS: This session will consist of general discussion of five topics suggested to me by the Proceedings so far conducted.

RANDOMIZATION OF THE FIRST PATIENT

In a study of prophylactic portacaval shunt surgery carried out in Boston, the first patient ever done at the Lemuel Shattuck Hospital was randomized [1]. In a study of another bypass operation, that of colon bypass for chronic encephalopathy in patients with cirrhosis, the first patient done by members of the Boston Interhospital Liver Group was randomized [2] but there had been three patients so treated before that time [3]. A few other examples of randomizing the first patient have been cited [4]. More recently the first patient to receive intraarterial streptokinase at Mount Sinai Hospital this last year was randomized [5], and as far as is known the first patient to receive atabrine for the treatment of HBsAg + chronic hepatitis was part of an RCT [6].

Received June 14, 1982; accepted July 1, 1982.

© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030285-14$02.75

FACTORIALLY STRUCTURED TREATMENTS

Exactly 30 years ago this fall we conducted a study in Kyoto, Japan [7], on the treatment of hepatitis, and we were presented with two very common treatments which needed to be evaluated. One was a high-protein, high-calorie diet and the other was bed rest. So we combined them in a two-by-two factorial design and found no interaction, and thus answered both questions with half the number of patients that we would have had to study if we had studied them sequentially.

We then did a second study because we had demonstrated an effect of the nutritious diet. This time we employed a two-by-two-by-two factorial design to test for the effects of calories, protein, and supplementary vitamins. We were able to segregate the effect as being one of the high protein in the diet and not of the calories or of the supplementary vitamins. Then half of each of the eight groups had early exercise and the other half stayed in the hospital for an extra ten days, and we demonstrated no bad effects of early exercise, making it a two-by-two by two-by-two factorial design. Again, four questions answered with minimal use of patient material.

It seems to me that the cardiovascular and the cancer clinical trial people are drastically underutilizing a very useful technique which could save a lot of time, effort and money. Certainly a good example would be AMIS [8] and BHAT [9]. Both of those studies could have been done at the same time with the same patients, changing the protocol for AMIS to start aspirin while the patient was in the hospital instead of three months after the MI. There was really no expectation that the two treatments would interact, and, therefore, the number of patients could have been half what it took to settle both questions. And, obviously, the cost saved for the government could have been enormous. I don't know how much BHAT cost, but the answer could have been available through AMIS.

DR. FRIEDEWALD: Eighteen million.

DR. CHALMERS: That would have been an $18 million saving if someone had thought in terms of a two-by-two factorial design.

In our collection of over 2,500 randomized control trials, less than 1% (we don't yet have the exact number) have factorially structured treatments, and two of them were designed and executed by me. So there are not many people who use the technique. I find the major problem lies with the biostatisticians who suffer from some kind of extreme bias against it. I have tried to sell the technique to a number of people on a number of occasions and they cite the problems of missing values and potential interaction as prohibitive. Yet it is the only way available to detect interaction, pharmacologically a most important phenomenon. So I don't accept that as a legitimate argument against factorial studies.

DR. MEINERT: I am always skeptical of free lunches. I don't think you get a free lunch with the factorially structured treatments because if you are really going to take advantage of the structures, you indeed have to meet certain assumptions and sometimes those assumptions are troublesome.

I agree with you that the structures may be advantageous but the "free lunch" you talked about with regard to the AMIS and BHAT may not be free. What would you do there? You'd have aspirin alone, aspirin and propranolol, and propranolol alone, and only one control group.

DR. CHALMERS: You would have to have equally sized groups to conveniently detect both main and interaction effects. If there is no aspirin–propranolol interaction and you have 1000 in each group, you then have 2000 who get propranolol and 2000 who don't to compare, and you have 2000 who get aspirin and 2000 who don't. That is under one assumption which is that you demonstrate there is no interaction, and I suspect that the numbers are large enough here to be able to demonstrate whether or not there is an interaction.

DR. FISHER: Well, there is less power for the interaction. If you take it to an extreme, take five drugs that you are pretty sure don't interact, you have 32 treatment groups.

DR. CHALMERS: We looked at protein, calorie, vitamin and early exercise effects in one study. A two-by-two-by-two-by-two factorial structure was used [7].

DR. FISHER: It gets harder and harder.

DR. CHALMERS: Because the numbers in each block get smaller and smaller. Yes.

DR. MEINERT: So, indeed, if you combine the treatments such as used in AMIS and BHAT into a single study you get by with one data center as opposed to two, certainly.

DR. CHALMERS: And you get by with 4000 patients rather than 8000.

DR. ELASHOFF: But if there is an interaction, then you don't get by with the 4000.

DR. CHALMERS: Right. And tell me some other way you could ever detect that interaction, which might be critically important if patients are going to go on aspirin and on beta blockers from now on.

DR. ELASHOFF: But you may not have the power to detect it because you only have 1000 in the control group.

DR. CHALMERS: But that is a pretty powerful study even so.

DR. MEINERT: But then it isn't so much an argument for factorially structured treatments as it is just an argument for doing larger studies.

DR. CHALMERS: No. Because if you got by powerwise, you would be answering two questions with 4000 patients instead of 8000. If it turns out there is interaction, then you would be learning something about the management of patients with regard to aspirin and propranolol which you never will be able to learn now because you never again will be able to have a control group taking neither aspirin nor propranolol. So, if you ever want to find out whether there is an interaction between aspirin or some other drug that might come along, you are stuck with the fact that you can't have a control group for propranolol and, therefore, you can't have a two-by-two factorial design.

DR. MEINERT: But the first thing you would like to do with that design is to increase the number of people in the control group. You have two treatment groups you are comparing.

DR. GEHAN: You could do that consistent with the factorial structures. You could have a factorial design with 2500 in each group.

DR. CHALMERS: All I would like to see is a little more evidence that biostatisticians have seriously considered factorially structured treatments. I was on the data monitoring committee for a study of hyperimmune gammaglobulin for hepatitis in which most of the patients were on steroids and I fought long and hard with the investigators, trying to persuade them that instead of withdrawing everybody from steroids or not allowing steroids (and the investigators were divided 50–50 as to whether steroids were essential or not) they should use a two-by-two factorial treatment structure and settle both questions; their biostatistician said we could not do that. It would confound things. It would confuse the clinical investigation. So they went ahead to compare a single treatment with the placebo. There was no sign of gross antibody effect in 120 patients [10]. And then they did a similar study of steroids that showed they also had no effect [11].

Patients are scarce for such trials. The investigators could have learned about both treatments simultaneously. Furthermore, there is a good reason for being interested in interaction when you are dealing with a gammaglobulin and a steroid. Both may have an immunologic effect on patients.

DR. MEINERT: Part of the problem has to do with the origin of a trial. It may be difficult to combine treatments into a factorial structure if the trials in question emerge from different initiatives. You should look sometimes at use of the structures in investigator-initiated trials as opposed to government-initiated trials.

DR. CHALMERS: The *N* might be too small.

DR. KLIMT: Tom, you will recall the rather unpleasant session at the NIH where it was suggested that one trial be carried out instead of three, to test aspirin, sulfinpyrazone, and persantine against placebo. So you could save on the size of the placebo-treated group and the amount of support structures. And you would have at the same time a comparison of three active treatments. For example, it is possible that the combination of propranolol and aspirin may well be the best combination we have for secondary prevention. But we have no such trial.

DR. CHALMERS: That is right. And you never can have it now.

DR. KLIMT: So I think that is the major reason for factorially structured treatments, not the saving.

DR. CHALMERS: I am not going to argue with you about the reason.

Another factorial study we did was an evaluation of ascorbic acid for both the prevention and treatment of the common cold [12]. We had a group getting no ascorbic acid either for treatment or for prevention; we had a group getting it only for prevention; a group getting it only for treatment; and a group getting three grams a day for prevention who got three more grams a day if they got a cold. It worked beautifully. We answered both questions simultaneously, except for the fact that some volunteers cheated and we had to stop the study because the placebo was detected and those receiving it dropped out more often than those on ascorbic acid.

DR. STAMLER: You would be interested to know that with Dick Remington's help—speaking about biostatisticians—a proposal for a partial factorially structured multiple risk intervention trial was made to NHLBI eight years before MRFIT started. That went down the pipe not on a scientific issue but the issue that Ted Cooper, as the new director, wasn't prepared to tolerate that kind of a grant in terms of the money involved. It had a sample size of 18,000. MRFIT went ahead with 12,000 for better or worse.

DR. MEINERT: You would also be interested to know that the Hypertension Prevention Trial, just recently funded, has factorially structured treatments. So there is no philosophical opposition to what you propose. And I am involved in that trial.

DR. GEHAN: In cancer clinical trials, I don't believe there is any bias on the part of biostatisticians against the use of two-by-two factorial designs. Among several examples that I am familiar with, I would like to mention an early study of the Brain Tumor Study Group in which a two-by-two factorial was used to study x-ray therapy and chemotherapy (BCNU) in the treatment of patients with anaplastic gliomas [13]. All patients received surgery and the two factors were x-ray and chemotherapy.

DR. CHALMERS: And some patients got both or neither?

DR. GEHAN: Right. This study established a definite benefit for radiotherapy in the enhancement of survival and a marginal benefit for chemotherapy. I think the problem in most cancer clinical trials now is that if you have two forms of treatment, you can get physicians to agree to give either or both of the two treatments, but it is difficult to get an agreement for the no treatment arm.

DR. SMITH: We are in the same boat now with multiple risk factor interventions in prevention of cardiovascular disease.

DR. CHALMERS: Incidentally, the Mount Sinai intra-arterial streptokinase study uses a two-by-two factorial treatment structure [5] in which the no treatment group stays on the Coronary Care Unit and gets standard therapy, and the catheterized patients receive nitroglycerin intra-arterially, streptokinase intra-arterially, or both. So it is a slight twist in the sense that one group doesn't go to the cath lab. But there is a group that gets both nitroglycerin and streptokinase, and one that gets one and one the other.

RCT CONTROL GROUPS AS A SOURCE OF UNNATURAL HISTORY

There has been some reference this morning to the control group of randomized control trials being a useful data base for the natural history of a disease. I brought along a slide from Thad Prout that is germane to this point [14]. It is a description of patient recruitment in a study of acute myocardial infarction (Table 1.) It documents the fact that a very small percentage of eligible patients may be randomized in a trial, and the patients studied are certainly not typical of patients with the disease. And Bill Friedewald told us this morning that only 25% of the BHAT patients seen by the investigators got into the study.

Table 1 Patient Recruitment in a Study of Acute Myocardial Infarction

% Remaining eligible from each protocol category		
% of the total	Protocol category	% of the previous
100	PTS admitted	100
65	Suspect AMI	65
52	Age eligible	80
19	Clinical MI	35
17	Alive 5 days	90
13	"Study" AMI	75
5	Exc. criteria	40
4	MD consent	85
3	PT consent and randomized	75

Source: Prout TE: Patient recruitment techniques in clinical trials. Controlled Clin Trials 1:313–318, 1981.

The point I am making is that I don't think clinical trials provide data that can ever be used to study the natural history of a disease process. I would call it the unnatural history of the disease when a placebo group in a randomized control trial is studied.

Another even better example probably is the Program on the Surgical Control of Hyperlipidemia. Henry Buchwald, what precentage of eligible post-myocardial infarction patients are randomized in your study?

DR. BUCHWALD: Ten percent of those that come into the clinic. About one percent of those that are screened by telephone. And about one per million of the overall adult population.

DR. CHALMERS: Those of you who know the POSCH protocol and know what a patient has to go through before he gets randomized, including cardiac catheterization, can appreciate the fact that this is an unusual group compared to all the patients who have had a myocardial infarction. It is a group that is very interested in doing something about their health and are intelligent and interested in research.

BIAS IN RANDOM AND NONRANDOM TREATMENT ASSIGNMENTS

There was a lot of discussion yesterday about randomization and the possible distortions that result from inadequacies in the process of assignment of patients to treatment groups. We have an elective course for medical students in which they attend sessions throughout the year and then work all summer, and have odd jobs after that working on various aspects of clinical trials. One of our students looked at the assignment processes in trials involving treatment of acute myocardial infarction, chosen because of three unique advantages of those trials [15]. One, there were a lot of them. In a 25-year period we found 138 trials of one kind of assignment or another. Two, the end point was constant and reliable in all of them. It was death in hospital. Three, most of the papers gave sufficient data on baseline variables to allow us to check on the validity of the treatment assignment.

We classified as BR, or blinded randomization, those papers in which the patient assignment was determined from opaque envelope, telephone call or prearranged bottle of medication.

Unblind randomization, UBR, included trials randomized from an open table of random numbers where the investigator may see the next number in line, by date of birth or entry into the study, by chart number or some other relatively random system by which physicians could know the treatment assignment when they select the patients. We could put coin flipping in this category because it is conceivable that the "wrong" coin could fall to the floor and require repeating because the flipper had a glimpse of the head or tail. Some of you might say that date of entry and chart number are not random, but in a sense it is chance that determines the order in which patients might come to the attention of the investigators; but they could be biased in the selection if they knew which treatment was next.

The third group is what we call the simultaneous selected controls. Patients in this group are assigned to treatment by a method more susceptible to clinical judgment than chance. The common example was the coronary care unit in which patients were admitted to the unit if a bed was available but not if it was not. One can imagine that a bed could quickly be freed up for a patient who really needed it. Or there are situations in which patients might be treated one way in one unit of a hospital and another way in another.

And, finally, the historical control, HC, in which patients were selected from hospital charts or from reports in the literature or from computerized data bases, sometimes matched by prognostic risk criteria, a process variably susceptible to bias.

In the blinded randomization group, the patients are selected and informed consent is obtained without any clue or knowledge of which treatment is coming up next. If the patients say yes, they are randomized, they get treatment A or treatment B; if they say no, they are not enrolled in the study.

In the unblinded randomization the randomization process has been executed before the patient is selected but the doctor may also know the result before the patient is accepted. That offers an enormous opportunity for bias in acting on criteria for acceptance or rejection, and bias may influence the vigor with which informed consent is sought.

All of you are, I am sure, aware that if you favor treatment A and you have a patient with a good prognosis and the patient is going to get treatment A, and the patient doesn't want to get in the study, you are willing to sit down and spend a good long time—I'm talking about your unconscious now—persuading the patient to enter the study. But if you favor treatment A and the patient looks pretty rocky, although he fills the written criteria of selection, and the patient says he'd rather not be in the study, you're much more easily persuaded. Doctors often don't appreciate that informed consent can be obtained from almost any patient at any time by any doctor if that doctor has a good relationship with the patient. I think the cause of recruitment difficulties is not with the patient at all; it is with the doctors.

DR. KLIMT: Tom, is it ever a practice to get informed consent after randomization?

DR. CHALMERS: Well, Marv Zelen, if he were here, would say it is the only way to do it in certain cases. You randomize patients first and if they fall in the group that is experimental, then you obtain informed consent [16]. I object strongly to Zelen's method of treatment assignment because for at least half the patients you know what the treatment is going to be when you try to get the informed consent, and you can be influenced by your bias.

DR. ROCKETTE: In NSABP Protocol B-06, where we encountered some difficulty with patient recruitment, we used a "prerandomization" scheme that differs from the one used by Zelen. One of the problems was that many of the physicians were reluctant to approach patients at the time of randomization about chance assignment to one therapy involving breast removal and another therapy involving cosmetic preservation of the breast.

Instead, if a patient is eligible for the trial the physician calls Headquarters and provides the patient's name and the eligibility criteria are checked. Then the patient is randomized and the physician informed what treatment she should receive.

Then the physician meets with the patient and informs her about the clinical trial and the various treatment options. The patient is informed as to her assignment. At this point she signs the informed consent. In doing this, all patients are followed, including those who refuse the protocol therapy. The effectiveness of the prerandomization scheme will depend on the number of women who refuse to participate once they are informed.

The number of refusals are closely monitored and if a physician has a high refusal rate, he is not allowed to continue participating in this trial. Our refusal rate on that protocol is about nine percent. We felt at the time that we had to make this compromise because with the usual randomization scheme we were not acquiring a sufficient number of patients.

I think this is an alternative where all patients give informed consent. My understanding of Zelen's was he had a standard treatment and those patients receiving the standard treatment did not have informed consent.

DR. CHALMERS: Yes, but if they got the experimental treatment they had to give informed consent.

DR. ROCKETTE: My understanding is that NCI considers this "qualified randomization" to be unethical. Secondly, in our protocol, we encounter the problem of determining the standard treatment. Some physicians would argue that it is the lumpectomy while others would consider it to be a total mastectomy.

DR. MEINERT: It sounds to me with prerandomization that physicians are likely to go to patients and say "we think this is the best treatment for you" and thereby not make them aware of the fact that they are involved in a randomized trial.

DR. ROCKETTE: Well, the informed consent that they sign spells out that they were randomized, lists all of the treatment options, and that they are part of a randomized trial.

DR. MEINERT: Then I don't understand the rationale for the prerandomization.

DR. ROCKETTE: The rationale is simply that the physician felt uncomfortable informing a patient of a chance selection to determine if the breast would be removed. Again, you can argue that you don't see the difference, but the physicians felt very strongly.

DR. FISHER: I would share Curt's concern, regardless of what was written on paper. It does take courage in a situation where a person is expecting you to be absolutely knowledgeable to say, I don't know. I have never heard of this design before.

DR. ROCKETTE: You are assuming that the patient just wouldn't bother reading the informed consent?

DR. FISHER: No, no! I am assuming that the situation is emotional. It is a combination of the written and verbal information you give. But the investigators have to say we don't know which treatment is best, unless they want to have a lot of dropouts after randomization.

DR. ROCKETTE: As I have already indicated, there can not be many dropouts if the procedure is to be effective. It's certainly not a "cure all" for a conventional randomization scheme.

DR. GEHAN: If Marv Zelen were here, I think what he would say is that unblinded randomization is very close to his randomized consent design. If you randomize for treatment A and seek consent for treatment A, that becomes essentially his randomized consent design if you accept the modification that there are no discarded patients. That is, if you seek consent for treatment A and the patient refuses it, Marv still would count him in the treatment A group and similarly for the treatment B group. So, if you eliminated that last possibility, it is essentially equivalent to his design.

DR. CHALMERS: In the case of controls that are selected from present or past patients (simultaneous or historical controls), one selects patients for an exciting new therapy and if they give informed consent, they get the experimental treatment, and if they say no, they are out. To achieve a treatment B group, you match or you look for patients by various methods susceptible to bias.

An additional point which I think people have lost track of is that there is no informed consent procedure involved in the selection of historical or selected controls. That to me is the most important thing of all. That is the primary reason, I think, for the frequent differences in the medically treated patients that are not part of a controlled trial. The patients we saw and considered for randomization but didn't get informed consent because they didn't look as though they were in good enough shape to go through surgery had a much worse prognosis than those who were enrolled [1]. When you have to explain to somebody exactly what is going on, you may eliminate a lot of critically ill patients.

When we looked at the results of the AMI studies classified by the four types of treatment assignment, we found that the blind randomization studies had the smallest number of significant differences in case fatality rates between treatment and control groups, and the historically and simultaneous controlled studies the most. The unblind randomization group was in-between. These differences were almost exactly parallelled by the distributions of pre-assignment risk factors. The blind randomization studies had the expected numbers of significant maldistributions ($p < 0.05\%$) and the others had progressively more, suggesting an increasing bias in case assignment.

TIME LAGS BETWEEN INNOVATION, RCT's, AND PRACTICE

I looked at the time lags between the first report of a new therapy (innovation), the first RCT (trial), and general adoption by practitioners (usage). The first reported series of ligation of the internal mammary artery we found was 1955 [17]. And then the two RCTs were performed in 1959 and 1960 [18,19]. The operation was promptly abandoned, but a possible factor was the advent of apparently better procedures.

The Vineberg operation, i.e., inserting the internal mammary artery into the myocardium, on the other hand, was first reported in 1957 [20]. And then the VA study which actually had been done quite a while before was only reported in the 1980s [21]. So the lag was 24 years. The first reported RCT for coronary bypass surgery was reported in 1977 [22]. So, in that case there was a minimum of 10 years before randomized control trials began to come out, with I guess the VA left main artery results being the first [23].

DR. FISHER: It turns out that the first patient treated with mammary ligation was in 1954.

DR. CHALMERS: Okay. Close enough. Obviously, ligation and the Vineberg operation are no longer carried out, in part because of lack of confirmation by an RCT but mostly because we have something better in saphenous vein bypass operations.

With regard to how often these operations are done, it is hard to get good data on doctors making decisions in individual cases. But what we have done is look at doctors' opinions about who should be operated on, as expressed in the published literature [24]. In this case we blinded the source of the paper and looked at the opinion and then we blinded the opinion and looked at the source of the paper, so that we could control our possible biases. There is a clearcut, highly statistically significant relationship in that whether or not doctors advocate coronary bypass surgery depends on how they are trained, charitably, or how they make their living, uncharitably.

The statement is true both for angina (chronic as well as unstable) and for acute myocardial infarction. The surgeons advocate surgery and the internists are not quite that enthusiastic.

With regard to radical mastectomy, the trials for radical versus simple mastectomy were in 1969 [25]—and 1973 [26]. Halsted introduced radical mastectomy in 1894 [27], McWhirter proposed use of simple mastectomy in 1948 [28]. And exactly the same thing holds here in that when we did a survey of opinions as to whether patients should have radical or simple mastectomy, we found that surgeons thought they should have radical mastectomy and the radiotherapists thought they should have simple mastectomy plus radiotherapy. Again, a statistically significant distinction.

In the case of prophylactic radiation after radical mastectomy, first proposed in 1936 [29], 23 years passed before the first trials, and all of them, with one minor exception, have been negative for survival [30]. The last survey I saw is that radiotherapy is done in a third of patients who are still having radical or modified radical resection. And, again, the radiotherapists say that in spite of the controlled trials you should give the patients the benefit of the doubt and give them x-ray therapy, and the nonradiotherapists say that because of the controlled trials the therapy is not indicated [24].

DR. FISHER: I think it is important to remember that the concept of randomization, which I view as one of the great intellectual contributions of the 20th century, really only came in in the 1930s with Fisher. So, in some sense with radical mastectomy, I think you might pick 1950 as a reasonable date when randomization was even a plausible alternative. Maybe over in England a few years earlier.

DR. CHALMERS: Stilbestrol for spontaneous abortion, which first was used in 1942 [31], was found to have no effect in five studies performed between 1950 and 1955. One was an RCT and the others used alternative cases or were assigned by chart number or some other nonblinded method. But three of the studies blinded the physicians and patients. Six historically controlled trials showed the drug to be beneficial. It wasn't until 1971, when cancer of the vagina in the female offspring was first described, that people started to diminish the use of stilbestrol [31].

We looked at all the textbooks of Ob-Gyn over this period of time and did find a spurt of enthusiasm around the 1950s based on the uncontrolled trials, but by about 1958 to 1960 most of the textbooks had come around to saying the drug should not be used. This is long before the 1971 cancer discovery.

DR. MEINERT: Tom, the date I assume you are using for assessment of time lapse is the date the trial was published, not when it was started.

DR. CHALMERS: These are all published dates.

DR. MEINERT: So they could have started several years earlier than the times you report?

DR. CHALMERS: Yes. Bed rest for hepatitis was first vigorously advocated in 1945 [32], shown to be ineffective in 1955 [7] and this was confirmed by another trial [33]. We did a survey in 1973 [31] which showed that in university hospitals physicians were still prescribing strict bed rest for about 50% of patients and in community hospitals in about 70% of the cases.

In the case of a bland or sippy diet for peptic ulcer, eight published trials have shown that the composition of the diet has no influence on healing, but it took 37 years for them to be done, and, again, a survey in 1973 showed that 35 of 38 doctors in university and community hospitals were still prescribing a bland diet and 15 were still ordering the sippy diet [31].

DR. STAMLER: When I was in medical school, Tom, I was taught by a man—this is one of the many things that got me interested in atherosclerosis—who reported seven or nine coronary deaths in Brooklyn. A strict clinical report, a clinical impression. He believed early coronary deaths were in people with peptic ulcers who had been drinking cream for years and were atherosclerotic.

DR. CHALMERS: Yes. And this, then, is a good example of an "it can't hurt" kind of therapy employed in spite of controlled trials showing it doesn't work, and now, in retrospect, it looks like it very well could have hurt.

DR. STAMLER: The same thing in diabetes, by the way, the diet for diabetes.

DR. CHALMERS: Yes. The unfinished saga of the oral diabetes drugs goes back to 1954 [34] when they were introduced, and 1970 [35], when the UGDP was published, a 16-year span. I didn't bring the data on sales but, as you know, they are finally coming down at a fairly steady rate, but it took about six or eight years after the UGDP report for usage to decline.

DR. MEINERT: I am told that trend is being reversed.

DR. CHALMERS: Well, the data I have are about two years old and it hasn't started up yet.

DR. KLIMT: The real question is whether industry can still get new patented hypoglycemic drugs on the market. The old ones don't bring much profit anymore.

DR. CHALMERS: So if one were to summarize the reasons why randomized control trials are not paid attention to as much as we would like, one reason is the "it can't hurt" philosophy, which I think is why stilbestrol continued to be used, and it could hurt. And also cream in the treatment of peptic ulcer. And another is the "have to do something" philosophy as in bed rest for hepatitis.

Another is the pressure from industry to sell drugs. I think the story of the UGDP is overwhelmingly impressive, and especially when one contrasts that study with that of Fries, the early VA study of antihypertensive agents [36], which according to our scoring system for randomized control trials is not as good a trial as the UGDP. It gets a lower quality score. It had lots of problems about selection of patients and soft end points; patients were removed if their blood pressure bounced up and it was not very double blind. And yet it was accepted universally and immediately acted on. And the UGDP, a better study, was violently fought for years, and the lack of action on mentioning the study in the package insert is a real black mark on the Food and Drug Administration.

DR. STAMLER: I think, Tom, you have to say there, too, that there was such a threat to a sector of specialization that members of that sector were encouraged to align themselves with industry. They were being told more or less at one and the same time that if you made a diagnosis of a so-called mild chemical, diabetes, the diagnosis and treatment might not even be proper. And they have now backed off that with new diagnostic criteria.

They were also told to use a diet that is the equivalent of Sippy, the diet carryover from the preinsulin era, low in carbohydrate, high in fat, high in animal fat, and then you dropped the caloric control when you got insulin. Now you are being told you developed a set of drugs, went out and detected all kinds of people to push those drugs on and that is disaster.

These three things tossed at diabetologists at once really were very threatening. So they were driven to an alliance, many of them, aside from other reasons, for their having an alliance with the drug industry. I know many of these people personally and I watched them in all of this. It was very interesting. Now many of them are having quieter second thoughts.

DR. CHALMERS: Yes. But it clearly was an element of commercialism not just in the advertising industry, not just in the pharmaceutical industry, but also in the clinic industry.

DR. KLIMT: And, thirdly, there was the convenience of the patient. So, against all three factors, how could you win, at least in the short term?

DR. STAMLER: That is right.

DR. CHALMERS: For those of you who are worried about the length of time it takes doctors to adopt the truth, my last point is that citrus fruits were first suggested in print for the treatment of scurvy in 1593 [37] and again in 1608 [38]—then in 1747 [39] came the clinical trial, which was not randomized but which was pretty conclusive—two patients got well on citrus fruit and the other ten, on five other treatments, got worse. It was 40 years before the British navy required lemons and limes and other citrus fruit on board ships at sea.

So, it took 154 years from the innovation to the trial, 40 years from the trial to practice.

REFERENCES

1. Garceau AJ, Donaldson RM, O'Hara ET, et al: A controlled trial of prophylactic portacaval-shunt surgery. N Engl J Med 270:496–500, 1964
2. Resnick RH, Ishihara A, Chalmers TC, Schimmel EM, and the Boston Inter-Hospital Liver Group: A controlled trial of colon bypass in chronic hepatic encephalopathy. Gastroenterology 54:1057–1069, 1968
3. McDermott WV Jr, Victor M, Point WW: Exclusion of the colon in the treatment of hepatic encephalopathy. N Engl J Med 267:850–854, 1962
4. Chalmers TC: Randomization of the first patient. Med Clin North Am 59:1035–1038, 1975

5. Rentrop KP: Randomized clinical trial of non-surgical reperfusion in acute myocardial infarction. (Study in progress)
6. Bodenheimer HC Jr, Schaffner F, Hirschman SZ, Goldberg JD, Chalmers TC: Randomized control trial of atabrine for the treatment of Hb$_s$Ag positive chronic hepatitis. (Unpublished)
7. Chalmers TC, Eckhardt RD, Reynolds WE, et al: The treatment of acute infectious hepatitis. Controlled studies of the effect of diet, rest, and physical reconditioning on the acute course of the disease and on the incidence of relapses and residual abnormalities. J Clin Invest 34:1163–1235, 1955
8. Aspirin Myocardial Infarction Study Research Group: A randomized, controlled trial of aspirin in persons recovered from myocardial infarction. J Am Med Assoc 243:661–669, 1980
9. Beta Blocker Heart Attack Trial Research Group: A randomized trial of propranolol in patients with acute myocardial infarction. J Am Med Assoc 247:1707–1714, 1982
10. Acute Hepatic Failure Study Group: Failure of specific immunotherapy in fulminant type B hepatitis. Ann Intern Med 86:272–277, 1977
11. Acute Hepatic Failure Study Group: A double blinded randomized trial of hydrocortisone in acute hepatic failure. Gastroenterology 76:1297, 1979
12. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM: Ascorbic acid for the common cold a prophylactic and therapeutic trial. J Am Med Assoc 231:1038–1042, 1975
13. Walker MD, Alexander E, Hunt WE, et al: Evaluation of BCNU and/or radiotherapy in the treatment of anaplastic gliomas. J Neurosurg 49:333–343, 1978
14. Prout TE: Patient recruitment techniques in clinical trials. Controlled Clin Trials 1:313–318, 1981
15. Celano P, Chalmers TC, Sacks H, Smith H Jr: Treatment assignment bias in controlled clinical trials. Controlled Clin Trials 2:76, 1981
16. Zelen M: A new design for randomized clinical trials. N Engl J Med 300:1242–1245, 1979
17. Battezzati M, Tagliaferro A, de Marchi G: La legatura delle due arterie mammarie interne nei disturbi di vascolarizzazione del miocardio: nota preventiva relativa di primi dati sperimentali e clinici. Minerva Med 46:1178–1188, 1955
18. Cobb LA, Thomas GI, Dillard DH, et al: An evaluation of internal-mammary-artery ligation by a double-blind technique. N Engl J Med 260:1115–1118, 1959
19. Dimond EG, Kittle CF, Crockett JE: Comparison of internal artery ligation and sham operation for angina pectoris. Am J Cardial 483–486, April 1960
20. Vineberg A, Walker J: Six months' to six years' experience with coronary artery insufficiency treated by internal mammary artery implantation. Am Heart J 54:851–862, 1957
21. Bhayana JN, Gage AA, Takaro T: Long term results on internal mammary artery implantation for coronary artery disease. A controlled trial by the participants, V.A. Cooperative Study Group. Ann Thoracic Surg 29:234–242, 1980
22. Murphy ML, Hultgren HN, Detre K, Thomsen J, Takaro T, and Participants of the Veterans Administration Cooperative Study: Treatment of chronic stable angina. N Engl J Med 297:621–627, 1977
23. Takaro T, Hultgren HN, Lipton MJ, et al: The VA cooperative randomized study of surgery for coronary arterial occlusive disease. II. Subgroup with significant left main lesions. Circulation 54: Suppl 3:III-107–III-117, 1976
24. Chalmers TC, Reitman D, Koffler G, Sacks H: Bias in the interpretation of clinical trials. Clin Res 30:549A, 1982
25. Kaae S, Johansen H: Simple mastectomy plus postoperative irradiation by the method of McWhirter for mammary carcinoma. Ann Surg 170:895–899, 1969

26. Roberts MM, Forrest APM, Blumgart LH, et al: Simple versus radical mastectomy. Preliminary report of the Cardiff Breast Trial. Lancet 1:1073–1076, 1973
27. Halsted WS: The results of operations for the cure of cancer of the breast performed at the John Hopkins Hospital from June, 1889, to January, 1894. Ann Surg 20:497–555, 1894
28. McWhirter R: The value of simple mastectomy and radiotherapy in the treatment of cancer of the breast. Br J Radiol 21:599–610, 1948
29. Portmann UV: Comparison of results in series of cases of carcinoma of the breast treated by postoperative roentgen therapy for prophylaxis with similar series in which operation was only treatment. Am J Cancer 27:1–25, 1936
30. Stjernsward J: Adjuvant radiotherapy trials in breast cancer. Cancer 39:2846–2867, 1977
31. Chalmers TC: The impact of controlled trials on the practice of medicine. Mount Sinai J Med 41:753–759, 1974
32. Barker MH, Capps RB, Allen FW: Acute infectious hepatitis in the Mediterranean Theater. J Am Med Assoc 128:997, 1945
33. Repsher LH, Freebern RK: Effects of early and vigorous exercise on recovery from infectious hepatitis. N Engl J Med 281:1393–1396, 1969
34. Franke H, Fuchs J: Ein neues antidiabetisches prinzip: Ergebnisse klinischer untersuchungen. Dtsch Med Wschr 80:1449–1452, 1955
35. The University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. Diabetes 19, Suppl 2, 1970.
36. Veterans Administration Cooperative Study Group on Antihypertensive Agents: Effects of treatment on morbidity in hypertension. J Am Med Assoc 202:1028–1034, 1967
37. Alberto S. (1593), cited by Lind J: *A Treatise on the Scurvy. 1973*. Birmingham: Classics of Medicine Library, 1980, p 317
38. Vizcaino S. (1608) cited by Lind J: *A Treatise on the Scurvy. 1973*. Birmingham: Classics of Medicine Library, 1980, p 332
39. Lind J.: *A Treatise on the Scurvy. 1973*. Birmingham: Classics of Medicine Library, 1980, pp 149–153

Conference on Recent History of Randomized Clinical Trials:

Concluding Session—General Discussion

Edited by Thomas C. Chalmers and Peter Amacher

DR. CHALMERS: Our goal in this last session is to have an informal discussion of several important issues raised by the preceding papers. It occurs to me that we should first discuss the distinctions between cancer and heart trials. Several of them were mentioned today.

The presence here of cardiovascular disease epidemiologists and cancer therapists certainly reflects a major distinction in the two types of trials. A lot of the heart trials deal with preventative treatments. Are there any randomized control trials for the prevention of cancer? No one has ever tried to randomize a MRFIT type trial with regard to cancer risk factors.

Another distinction is prerandomization stratification. There are real differences between the two fields in this regard. Don't the oncologists stratify more often than the heart disease types?

DR. GEHAN: I think that in cancer the stratification is more according to prognostic features in the patients and less on clinics; my interpretation of the discussion at this meeting was that in heart studies the focus is more on clinics and less on prognostic features.

DR. SMITH: Isn't it a function of numbers, too? I wonder if we are not more comfortable allowing randomization to suffice for us in the cardiovascular trials since we are dealing with larger numbers and needn't feel so compelled to stratify.

DR. FISHER: I think most therapeutic trials in cardiovascular disease, unless you had a trial in congestive heart failure, do tend to be larger than the typical cancer trial. It is a function of the observed mortality rates, and in cancer there is a very pessimistic prognosis in all sorts of subsets of patients by cardiovascular standards. It is a statistical issue. You need end points.

DR. GEHAN: My list of differences started with diseases. Cancer being many diseases, there are many hypotheses being tested; in the heart area, at least the things that we heard today, there are rather few areas of study.

DR. CHALMERS: Recruitment is certainly different when you are dealing with sick patients and you can only get your subjects from practicing doctors than it is when you are dealing with normal people whom you can recruit from the population.

DR. STAMLER: The recent trials that have started early after MI cannot recruit publicly. It's a very interesting difference. You have to get patients out of coronary care units.

Controlled Clinical Trials 3:299–309 (1982)
© Elsevier Science Publishing Co., Inc. 1982
52 Vanderbilt Ave., New York, New York 10017

0197-2456/82/030299-11$02.75

DR. MEINERT: The coronary artery surgery study didn't recruit from the public, so it depends on the study.

DR. FRIEDEWALD: I would argue even in those, Jerry, you don't ask the physician to refer patients. You get his permission and you go to CCU on a regular basis.

DR. FISHER: I thought Dr. Holland said yesterday that he thought that every tumor site was a different entity in some sense, and might respond to different chemotherapy. And there were literally hundreds of them. So it is not just a question of recruiting the total number of patients. In cardiovascular disease there certainly are many different entities, particularly if you get into congenital heart disease and so on, but basically for most of the trials the underlying idea is probably a single etiology of coronary artery disease. There are a lot of theories about what it is about, and if somebody can prove one they have a Nobel Prize. But my understanding of cancer is that it is not just a question of recruiting a lot of people, but you can't logically lump lung and breast tumors together in one treatment protocol and think you are going to have the same effect. Is that correct?

DR. ZUBROD: I think the example Dr. Holland used was acute lymphocytic leukemia, which, from the pathologist's viewpoint, is more or less a single disease. But because some of these patients have been cured by therapy and others not, a further look at heterogeneity was made. With the availability of surface antigens you can now demonstrate—Jim said at least three—perhaps as many as ten different subsets within the morphologic entity of acute lymphocytic leukemia. You have to deal with them as separate diseases. They respond to different drugs and have a different natural history.

DR. FISHER: I submit one difference from a statistician data handler's point of view. In cancer you have sufficient different combinations of therapies to have ongoing funded data centers for the different oncology groups that have a long history, 25 years or whatever, where they have many, many different treatment protocols underway at once. In heart disease you have huge mega-trials by the NCI standards where you can spend enough money to fund several hundred cancer trials. So there is a tremendous difference in scale and the way in which the trials are organized and how you recruit people.

DR. CHALMERS: But you are talking about preventive heart trials and therapeutic cancer trials.

DR. FISHER: No. I am talking about therapeutic trials such as the Coronary Artery Surgery Study. And 780 (the sample size of the study) would be an extremely large trial in the cancer field.

DR. CHALMERS: That is still a trial of prevention I would say. When you are dealing with treatment of acute myocardial infarction, with a base-line mortality of 15%, you don't need that many.

DR. ROCKETTE: But I thought you were talking about what tends to be a difference. In other words, I think in the cardiovascular area it is how fast the end point occurs. In fact that is true. You mentioned some cardiovascular diseases like MI which have a high event rate. On the other hand, in breast cancer, which tends to have a better prognosis than many of the cancers, we have to do relatively larger trials as in the cardiovascular disease field.

DR. CHALMERS: But that is again prevention. You are preventing recurrence.

DR. ROCKETTE: Yes, I guess. Some people might argue that what you are treating is already there but subclinical. The question in cancer is how rapidly does recurrence occur. The sample size estimates are really based on that.

DR. FISHER: From the self-interest point of view of somebody who directs a data center, continuity of personnel is crucial. It is nice to look at someone like Marv Zelen where there is a certain assurance of continuity, rather than saying that when this trial ends, if you want to stay in heart disease, unless a big new trial comes along you disband, or you prematurely take on additional work, in which case you do a mediocre job on both. I don't know what the best model is. There is a lot to be said for experienced investigators. On the other hand, there is a lot to be said for new blood and enthusiasm and new Request for Proposals. But from an operational point of view, this to my mind is a very big difference between the cancer and heart institutes.

DR. CHALMERS: One of the major problems, as I see it, that distinguishes the two is that quite often in cancer clinical therapy trials you can't tell when patients get sick whether they are sick because of the drug you gave them or because of their disease. Yet you still have a fairly high rate of deaths in the placebo or the standard therapy.

DR. GEHAN: It is very rare to have placebo or double-blind studies in cancer.

DR. CHALMERS: But you could have a double-blind study much more often by using a disguised therapy so that if you have two cancer patients getting different therapies being compared, each of them should be getting a placebo for the other therapy. But it is not done as often by cancer people as it should be.

DR. GEHAN: I think it is fairly well accepted that some therapy is good and therefore physicians are going to give treatment rather than a placebo or no therapy. They don't do double blind because they know something about the side effects. It is not double blind if people start losing their hair.

DR. GREEN: Randomized cancer studies involving placebo do exist. Use of placebo depends on the particular cancer site and the specific stage of disease. Examples are estrogen versus placebo for prostatic cancer, and levamisole versus placebo in double-blind immunotherapy trials.

DR. FISHER: Bill Friedewald has had about as much experience in trials in the cardiovascular area as anybody here. I wonder what he thinks about the importance of having experienced investigators in a trial or having a group who have been through the mill and who know the problems of randomization etc. Is the mix of new and old investigators important?

DR. FRIEDEWALD: As I had on one of my slides, I think it is. The new people are enthusiastic and there is a learning process involved that is healthy. A lot of questions are asked of the more seasoned investigators.

DR. MEINERT: A difference I see has to do with the organizational structure of a multicenter trial. In heart disease and other related fields it is standard operating procedure to have a data monitoring committee to watch for emerging treatment effects which require alteration of the study. Results of the trial are not presented outside that committee until there is something to report. The data monitoring function in many cancer trials is less formalized. Data are presented and published as the trial proceeds.

DR. ROCKETTE: We don't have a data monitoring committee in our study. We generate our report in-house every six months. We do not present all the results. For instance, protocol six has never been presented to anybody. Our first step, and I guess it serves the role of a data monitoring committee, is to present results to our executive committee. The executive committee for the most part consists of the study clinicians.

DR. CHALMERS: The process is entirely different from that of the typical multicenter heart trial.

DR. SMITH: I recently developed a concern that in the multistudy centers, cardiovascular centers, that there is a problem because the staff is very often interchangeable and the protocols are very different. At least they were written differently. It becomes very difficult for that staff to maintain those differences. I am wondering if that is a problem in cancer chemotherapy, for example. Are the similarities greater than the differences in protocol?

DR. ZUBROD: I think the similarities are greater than the differences. There are three different phases to cancer protocols. The first is an early work-up. In the second phase we try to get some evidence of therapeutic efficacy. In phase three we are concerned with a comparison of new and standard treatments. Different institutions will involve different phase mixes. But on the whole, I think there are great similarities in protocols and the data handlers are quite accustomed to handling things more or less in the same format.

DR. GEHAN: I think that size is a problem. In the Southwest Oncology Group, when we go to a meeting, we are reporting on something like 40 to 50 different protocols. Having all of these committees you speak of is impossible. So the statistical center and the chairman of the study are the key people involved in monitoring that particular study. I think in cancer one of the big differences is that ongoing results are presented, often in coded fashion, treatment A versus treatment B; you don't know what the treatments are.

DR. CHALMERS: Isn't there an ethical problem when you go back to investigators and say the p value is 0.06 and we need only a few more patients to prove that this drug is less lethal than the other. Doesn't this present an ethical problem to the investigators who have to say to their patients that they would like permission to randomize them to drug A or B because they have a study going and they don't know which treatment is better, when in fact they do know which treatment is better, except that the p value is 0.06 instead of 0.05.

DR. ZUBROD: It certainly does present an ethical problem, and a difficult one to resolve.

DR. CHALMERS: The Heart Institute has resolved it by having an independent data monitoring committee to review interim results. The patients are told as part of the informed consent process that the doctor doesn't know which drug is better, and the data are being monitored by an expert committee.

DR. ZUBROD: You must realize though that the number of diseases, the number of investigators and the number of protocols have made for a decentralized structure for the whole cancer operation. Whether you can impose a single data monitoring committee somewhere centrally to cover this I would doubt. But somehow in cancer activities we ought to be taking a lesson from the cardiovascular studies and find some way of including a relatively impersonal data monitoring group. It may be that Phase I trials could have a central policy advisory group or data monitoring group. It would help with the ethical dilemmas and with public image.

DR. GEHAN: As progress is made in cancer you are making smaller leaps. There isn't that much of an ethical problem. It is very rare now that you do a study where one treatment is tremendously better than the other. They are a little bit better than the other.

DR. STAMLER: I wish to make a comment about an aspect which seems to be taken for granted here, but based on experience I think needs to be discussed. That is that by setting up an outside review group of quality that one assures a top level review. I don't want to debate here the issue of an internal group versus an outside group. That is not the point I am trying to make. My experience, in both the Hypertension Detection and Follow-Up Program and the Multi-Risk Factor Intervention Trial with an outside review group was that an outside review group did not by any means assure a level of review that made it possible to say "everything is in good hands, you don't have anything to worry about." It was our experience that the investigator group which is closer to the work, closer to the data, more wedded to the work, if you will, can play an important role in seeing to it that the data and safety monitoring committee look at the data in all the ways that they need to be looked at. Bill, is that an unfair statement?

DR. FRIEDEWALD: No.

DR. STAMLER: I think it is very important now to make that heart model seem as if it is a beautiful, perfect model that works magnificently. The fact is that the people who sit on those boards are very busy people, very qualified, very prestigious, and part of that is they all have too much to do. As a result they are likely to come in the night before, look at the tables, and think off the top of their heads. And that is a serious problem.

I don't know how anyone can sit down—speaking from personal experience—with the protocols from 40 studies and in a reasonable period of time digest the data from them. I would love to see the process in action. I find the prospect mind-boggling. Whoever sits down with data reports on 40 centers in a finite period of time and says this is our six monthly meeting, we will review the results of 40 protocols and come up with mature, rational, scientific decisions?

DR. GEHAN: Nobody does that.

DR. STAMLER: What do you do then?

DR. GEHAN: Well, you have certain physicians who work mainly in lymphoma consider the lymphoma studies. They have about six studies to consider. And then somebody is in myeloma. They have another five or six. In leukemia they have ten or twelve studies. The sarcoma committee has some, and lung cancer has some. So the physicians are specialists. They are considering only the studies that are in their specialty area.

DR. STAMLER: Ten or twelve at a fling, from my experience is quite a chore. Even five or six.

DR. GEHAN: But it is a spectrum of studies; Phase I, Phase II, and Phase

III. Some of the studies are already closed, some of them are being closed. There are not that many that require decisions. They are either in an early stage, too early to make a decision, or the decision has already been made.

DR. GREEN: Considering those trials on which you are ready to come to a decision, I haven't heard anything to suggest why an outside group would be any better able to make that decision than the investigators themselves.

DR. CHALMERS: There is an ethical reason. The investigators can't continue to enroll patients when they just need one more pair to prove—

DR. MEINERT: Tom, you are assuming that these decisions are always made while recruitment goes on. A lot of the decisions come long after recruitment is finished.

DR. CHALMERS: Then an investigator is faced with a patient coming saying, "I'd like to get out of the protocol and get that other drug."

DR. STAMLER: Let me be a devil's advocate for a moment. Why isn't that ethical "monkey" on the back of that policy board just as much as the investigators? I don't understand the logic of that.

DR. CHALMERS: Jerry, you have to start with the logic that all trials are unethical once the first pair of patients have been studied. You have to accept that and recognize that trials are better than giving a new drug outside of a trial. Then you say how can you make it easier for people who are taking care of patients and how can we make it easier for the patients who are involved in trials not to be asking constantly whether they are continuing on a drug that is falling behind. The constant switching that would result is both bad therapy and bad research. And bad research is unethical!

DR. STAMLER: Please note that this doesn't make the issue of external versus internal a fundamental issue. The CDP solved that problem by having a data monitoring group that was part internal, part external. But until we brought Ken Berge in there, none of us were seeing patients.

DR. MEINERT: I do agree with what Jerry says with regard to the quality of committee work. My own view is that a data monitoring committee functions better if some of its members are from the study.

DR. STAMLER: I have yet to hear a good justification for the MRFIT and HDFP model of formulating a data monitoring committee versus the CDP model.

DR. FRIEDEWALD: There is a difference in how the committees are perceived from the outside. We could go before the ethical review committee for HHS (Health and Human Services) which I had to do twice and say the committees are totally external to the study. The internal members who represent the investigators but don't see patients are on the committee but do not have a vote.

DR. STAMLER: I'm not concerned with voting. If you have markedly split votes in such situations you are in deep trouble.

DR. FRIEDEWALD: The key principle is that no one directly benefiting from a vote or in fact seeing patients should vote. I wouldn't argue that the experience in the CDP was in any better or worse than for MRFIT or HDFP. As a matter of fact, in many ways it was better, because the committee was so large.

DR. MEINERT: If I was a patient in a study I would much rather be one in a study where I knew the data monitoring committee contained people from in the study and familiar with the details of the protocol, rather than—

DR. CHALMERS: Don't forget the biostatistics center is there, as part of the general make-up. It is not totally outside.

DR. MEINERT: Staff of the biostatistics center do not understand all the data. They need help from clinicians.

DR. FISHER: Tom, I would disagree with your assessment that a trial is in a sense unethical after the first pair of patients have been enrolled. This brings up a very deep and serious question that does not get adequate debate, and maybe for good reason. How much risk can the individual patient take implicitly when they start in the trial and as it goes on. Let's take, for example, a trial of surgery in the early days, say heart valve surgery, where you have a high mortality. I don't care if you call them treatment A and treatment B. If you present survival curves, everybody is going to know which is which because of the high surgical mortality. Everybody understands it and accepts it. Even if there was a statistically significant difference, nobody would be in favor of stopping the trial. Nobody in their right mind would use a proportional hazards model and the various statistics associated with it. So you don't just draw logarithmic curves—I realize I am getting a little technical here.

It is a situation where you cannot effectively blind the data, no matter how you go about it. I think the policy board or data monitoring committee has as severe an ethical problem as the investigator, but there may be a lot to be said for having it discussed in a context which does not botch up the science of the trial. By the way, I think it is ethical for people to commit themselves to a certain amount of risk.

DR. ROCKETTE: Why are you having data monitoring and policy boards? Is this just a game you play to cover yourself, especially if you really believe a trial is unethical after the first pair of patients have been enrolled? Is that the point you are making, that we should have this board to cover us?

DR. CHALMERS: No, it is not a game we are playing to cover ourselves. It is a poor study, or no study at all, if patients are dropping in and out with regard to treatment, depending on data trends that are often meaningless. One can expect physicians to be swayed by emerging data trends since they are personally involved in the trial. The analysis job requires thought and discussion by experts. In the multicenter cardiovascular trials this job is given to a single monitoring committee, rather than to individual patients and their doctor.

DR. MEINERT: But I am not willing to buy the notion of bias. The possibility of bias depends on the nature of the outcome measure. It is not too easy to bias the measurement of death. There may be bias in an unmasked trial in the way treatment is applied. That may be a problem. But I don't think it follows immediately that if a trial is unmasked it is by definition biased. I think you have to look at the situation to determine the chance of serious bias.

DR. FISHER: How do we balance shortterm mortality and longterm mortality? For example, doesn't it ever occur in cancer that somebody comes up with a treatment that everybody realizes will cause some early deaths in the hope of extending life in some survivors?

DR. SANCHO-GARNIER: Yes. That is true for chemotherapy.

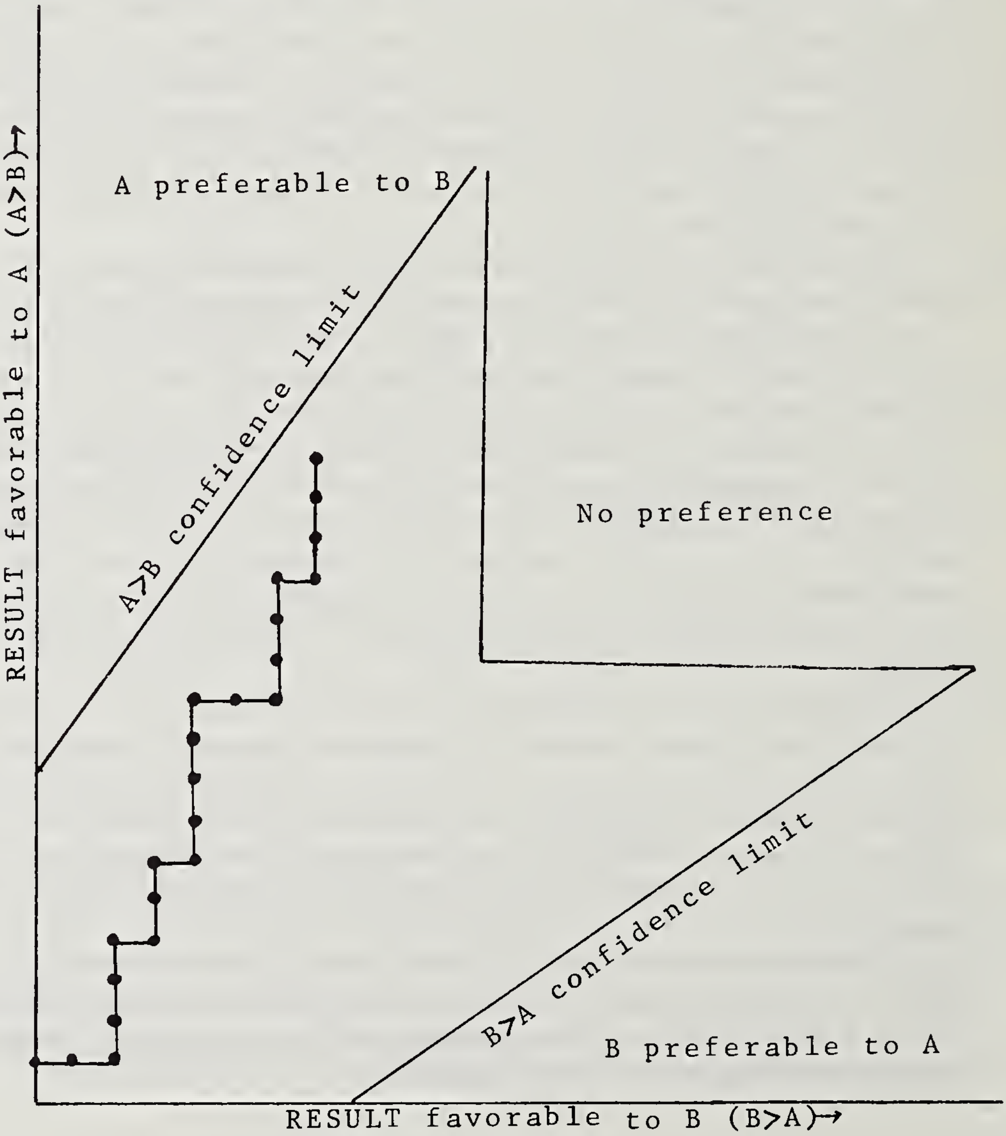
DR. FISHER: How do you blind results in that case?

DR. GEHAN: The identity of the survival curves is masked. The toxicity is not.

DR. ROCKETTE: Toxicity depends on the treatment. We have had two toxicity deaths in 6000 or 7000 patients. So, again, it gets to the specific disease you are talking about. If you remember some of the curves I presented, for breast cancer, you are dealing with few relative cancer recurrences.

DR. CHALMERS: I still maintain that if the practicing physician knows that the p is less than 0.06, but not less than 0.05, for A being better than B, he or she will be hard put to continue randomizing or to keep patients on the "worst" therapy. The dilemma is presented in Figure 1. Using sequential

Figure 1. Sequential analysis of randomized clinical trials. Chalmers, TC, and discussants: How to turn off an experiment. In: *Ethical Safeguards in Research on Humans,* Cooper JD, Ley HL, (Eds.) Interdisciplinary Communications, Associates: Maryland, 1976, vol 5, pp 119–143, with permission.



design as a model, at the last stages of such a study, A is probably better than B, could still equal B, but could almost never be worse than B. So what do you tell the patient? I would demand A because it is either equal to or better than B. What is the critical probability level at which the individual physician will abandon randomization and opt for one treatment or the other: at $p < 0.10$, $p < 0.20$ or $p < 0.49$?

DR. ROCKETTE: Why would you show patients data after each pair of patients is enrolled? Wouldn't they opt for the treatment that is ahead at that time?

DR. CHALMERS: That's right. That's the point I am making! And the patients who have volunteered for the study have been betrayed.

DR. GREEN: But, Tom, if you talk that way and believe that it is unethical for the physician to randomize a patient in the face of a statistically insignificant trend, however small, then if you are a member of an external committee you should feel the same way. I do not agree with the ethical distinction between the treating physician and the external committee. The justification for continuing properly randomized trials, which applies to both the physician and the committee, derives from the fact that it is not more ethical to select therapies based on random or chance occurrences. A decision is made at the beginning of the trial regarding what degree of evidence will be considered as unlikely due to chance alone. When such a point is reached, the trial is stopped. Furthermore, in cancer trials it may be particularly important not to reach a premature conclusion based on chance results in favor of a regimen, if that regimen produces serious toxicity, until there is reasonably convincing evidence that patients may benefit from receiving it.

DR. CHALMERS: I still think the only way you can solve the problem is have experts making the decisions who don't have to face patients. Those patients have gone to their doctor and have said, "I am in your hands, Doctor. One of the conditions of my coming to you, by tradition of the practice of medicine, is that you are going to do your best for *me*." Now I think you can do that as part of a randomized controlled trial if, in the informed consent, there is an understanding between you and the patient that that part of your arrangement is deferred by you to somebody else. You know, and the patient should understand, that if you knew the trends, even though they might be due to chance, you wouldn't be able to act normally in dealing with the patient, and that would be the end of the study.

DR. BUCHWALD: I don't think the investigators would act differently in that role. Knowledge doesn't necessarily determine change in action. And there is another ethic that you haven't even discussed, the ethic of wanting not to live with perpetual ignorance. We have a responsibility for more than the individual patient.

DR. CHALMERS: But most patients understand that when you are faced with a critical decision having to do with their health you are going to consider first their personal safety and welfare, and second, the search for truth via the study.

DR. BUCHWALD: Words get loaded.

DR. MEINERT: Why do you think patients would drop out of the study if given the results of the trial, Tom? What is wrong with presenting the data. Shortterm beneficial effects may be offset later by adverse effects. That was a concern in the Diabetic Retinopathy Study. The study continued for some time after emergence of a beneficial effect to try to help make certain these were not offset by adverse effects that occurred later.

DR. CHALMERS: I would say from the example (Fig. 1) that the probability of B being better than A is so slim that if I saw these data I would say thank you very much, take me out of the study and put me on A, put me on the treatment that is ahead, because it can only be equal to, never worse than B.

DR. FISHER: It depends on how you construct your boundaries. A boundary crossing early in the study may not be as important as you think.

DR. MEINERT: After listening to this discussion I am getting to the point of thinking one ought to show the patients the results as we go and live with the consequences. If you can't get any more patients to enroll, so be it.

DR. FISHER: We are all ethicists. That is a moral imperative in life. You have to make ethical decisions. I say Tom, I hope with a smile, that your ethical position is not defensible. If it is unethical for the person treating a patient if he saw the study data, it is unethical for the external policy board to allow treatment.

I do think people can take a certain amount of risk. In my own life in a lot of situations I would be willing to take some risk for the betterment of humankind.

DR. CHALMERS: You can say that now, but when you have a tumor growing and you are beginning to plan for death or survival I think the desire to die for mankind becomes diminished.

DR. FISHER: I said a certain amount of risk.

DR. SANCHO-GARNIER: I think when you accept participating in a clinical trial, you know that the ethic of the clinical trial is a collective ethic and you accept this.

DR. CHALMERS: I think it is conceivable that a fair number of sick people don't want to be involved in the decisions, want some reassurance that the doctor is thinking of them first and not thinking of other people and not thinking of ignorance or knowledge. The great majority of people want their doctor to think about getting them well when they are sick.

DR. MEINERT: Why shouldn't we ask patients at the outset do you want to see the data as they are emerging or do you want to be masked? Why not give them that choice?

DR. CHALMERS: You could. Except that I think if you showed the patient the data as they were emerging the patient would opt out pretty soon. Then you really would have no scientifically valid or ethical study. That's just an estimate, I don't know that.

DR. MEINERT: It's like the estimates we had that we'd never get any patients if we had to go through the informed consent process in the 1960s.

DR. GEHAN: There is one other difference between cancer and heart disease I wanted to bring up. In cancer, the therapy moves in sequence from Phase I to Phase II to Phase III. There's got to be some kind of sequence for heart but it hasn't come out in our discussion.

DR. BUCHWALD: The heart field started out with all that wonderful epidemiology and everybody believed it all—and I think some of us do believe it still. Then the idea was to use this powerful tool, the randomized clinical trial, and just put the final dot on it. And then the frustration started. As I said in my presentation on the oldest field in cardiac trials, lipid intervention, it was so clearcut—Ansel Keys did his studies and Jerry did his studies, all the great lipid epidemiologists of the last 40 or 50 years contributed. It was clean, absolutely clean. Randomized clinical trials were supposedly going to frame it and put it on ice forever. And they started out to do just that and they have kept at it but, all of a sudden, they can't put that frame around it. That is what has been difficult.

DR. GEHAN: So your Phase I and Phase II is epidemiology.

DR. MEINERT: No, no. That's not true. For the drugs you are testing there are indeed a Phase I and a Phase II, but they are done by the drug company or somebody else, whereas you are doing the Phase III.

DR. SMITH: I think that one of the differences that is getting in our way is that most of us are not involved in Phase I or Phase II cardiovascular drug evaluations. You, the cancer people, are involved in a discovery phase, if you will. We in the heart field are in an evaluation phase. We are in Phase III trials most of the time. You are in Phase I or II trials. That is a big difference.

DR. GEHAN: There are a tremendous number of chemotherapy people that have really very little interest in Phase III trials because they think they know the answers based on having Phase I and II trials.

DR. CHALMERS: And yet Phase III studies have often failed to confirm the enthusiasm that had grown out of Phase II studies. The final comparative trial is still essential before the new therapy is marketed.

This has been a frank and useful discussion.

Can any recruiting firm work "wonders" for your career?

No! There is no magic to advancing a career. There is a plan, however, and there are techniques.

Our plan will assist you at no cost. Research and quantitative scientists have relied on us for more than a decade as their "eyes and ears" for new positions nationwide.

Our techniques are grounded in high standards we've set for ourselves. Candidates, as well as employers, tell us they appreciate our distinctly low-key manner and the highly professional way we handle information. Employers know we carefully screen qualifications before sending them resumes; candidates know we actually compare their requirements and backgrounds with each new position we uncover in their field.

There's no magic about it. You've built your reputation on hard work, intellect, and dedication. We've done the same. Our reputation has grown in the scientific community and, because of this, our business has prospered while helping many professionals advance their careers.

Send a copy of your resume or c.v., along with a note describing your parameters in terms of type of position, salary, geographic location, etc. There will never be any charge to you, and you'll never be obligated to take any action whatsoever. But we will start working for you right away, and we won't stop until you tell us to. We're experienced career counselors, so you may very well advance soon—but *not* by magic.

More than a decade of building successful careers for toxicologists, pharmacologists, physicians, veterinary pathologists, C.R.A.'s, medical writers, statisticians, and scientific computing and regulatory affairs specialists.

Robert Wilkinson Associates

Box 300 ● 500 Bradley Hill Road ● Blauvelt, N Y 10913 ● 914-358-7200

J.B. CHASSAN CONSULTATIONS

763 Bloomfield Avenue/Montclair, New Jersey 07042/(201) 746-9555

Biostatistical Services Include:

NDA PREPARATION
MULTI-CENTER STUDIES
EFFICIENT SMALL SAMPLES DESIGNS
INDIVIDUALIZED, CLEARLY WRITTEN REPORTS

Elsevier Biomedical proudly announces:

ONE journal for procedures, commentaries, reviews, case studies, editorials and more–

DIAGNOSTIC MICROBIOLOGY AND INFECTIOUS DISEASE

Now *one* journal will keep you in touch with the latest developments—procedures, unusual cases, important books, controversial issues—in clinical microbiology and the diagnosis and treatment of infectious disease. Headed by a distinguished editorial board, every issue of ***Diagnostic Microbiology and Infectious Disease*** will bring you: full-length articles, critical reviews, case studies, book reviews, and editorials in • bacteriology • immunology • immunoserology • infectious disease • mycology • parasitology • virology

Editor:
William J. Martin, Ph.D.

Associate Editors:
Norman L. Goodman, Ph.D.
Sherwood Gorbach, M.D.
John M. Matsen, M.D.
Thomas F. Smith, Ph.D.

First Issue: January 1983

Special Offer to Charter Subscribers!

Individuals subscribing before November 15, 1982 can do so at the special rate of $32.00, a discount of 20% off the regular $40.00 subscription price. Enter your subscription to this major new journal today and save as a charter subscriber!

1983 Subscription Information

Volume 1 (4 issues) • Institutional Rate: $80.00 • Personal Rate: $40.00
Special Charter Subscriber Rate: $32.00 (Deadline: November 15, 1982)

For air delivery in the U.S. and Canada, add $12.00. Subscribers outside the U.S. should add: $10.00 for surface mail; $20.00 for air delivery to Europe and Mexico; and $26.00 for air delivery to the rest of the world.

Remember, the deadline for charter subscriptions is November 15, 1982, so act now!

PLUS–A Special Offer to Charter Subscribers!

ORDER FORM

Please enter my subscription to ***Diagnostic Microbiology and Infectious Disease,*** Volume 1 (4 issues), 1983, at the following rate:

☐ Special Charter Subscriber Rate: $32.00
(Deadline: November 15, 1982)

☐ Regular Personal Rate: $40.00

☐ Institutional Rate: $80.00

For air delivery in the U.S. and Canada, add $12.00. Subscribers outside the U.S. should add: $10.00 for surface mail; $20.00 for air delivery to Europe and Mexico; and $26.00 for air delivery to the rest of the world.

Return to:
Journals Fulfillment Dept.
Elsevier Science Publishing Co., Inc.
52 Vanderbilt, Avenue, New York, N.Y. 10017

Payment

☐ Check enclosed.
☐ Charge to my credit card:
  ☐ AMEX
  ☐ VISA
  ☐ Master Card (issuing bank # ______________)

Account # ______________ Expires __________

Signature ______________

Name ______________

Address ______________

____

City ______________ State__________ Zip__________

Equip yourself with these major references in clinical chemistry...

A clear, compact, and comprehensive guide to—

ESSENTIALS OF CLINICAL BIOCHEMISTRY

D.N. Baron, Royal Free Hospital School of Medicine, London, U.K.

Essentials of Clinical Biochemistry is your handbook to this fast-growing field of laboratory medicine. Opening with chapters on the principles of laboratory diagnosis (with an important section on test sensitivity) and general metabolism, it proceeds with discussion of abnormal metabolism of single or related substances, and the special clinical biochemistry of disorders of organ function.

Contents

Introduction. General Metabolism. Water and Electrolytes. Acid-Base Regulation. Carbohydrates. Lipids. Nitrogenous Compounds. Plasma Proteins. The Haemopoietic System. The Endocrine Glands. Calcium, Phosphorus, and the Bones. The Liver. The Kidneys. The Alimentary Tract. The Nervous System. Appendices. Index.

1982 304 pages 0-444-00684-2 $26.95

(Distributed outside North America by Hodder & Stoughton Ltd., United Kingdom)

A "core book" selection by the American College of Physicians—

CHEMICAL DIAGNOSIS OF DISEASE

Edited by **Stanley S. Brown** and **Frederick L. Mitchell,** MRC Clinical Research Centre, Harrow, U.K.; and **Donald S. Young,** Mayo Clinic, Rochester, Minnesota

"...a thick and impressively detailed volume that will serve both clinicians and laboratory scientists as a comprehensive and authoritative guide to modern laboratory medicine... of great assistance to all physicians who seek more efficient and cost-effective use of the laboratory."

—New England Journal of Medicine

"...a comprehensive treatise on the meaning of clinical chemical tests... Everything is documented, and the references are cited with complete titles... Clinical chemists will find this book a valuable guide... It appears to be the sort of book that will see heavy use as a reference work."

—Journal of the American Chemical Society

Contents

Biological Variability. Chemical Analysis of Body Fluids Other Than Blood. Hydrogen Ions and Blood Gases. Glucose Metabolism. Lipids. Proteins. Enzymes. Calcium and Magnesium. Renal Function. Liver Function. Gastric Function. Intestinal and Pancreatic Function. Thyroid Function. Steroid Hormones. Fertility, Pregnancy and Contraception. Haematological Biochemistry. Abnormal Haemoglobins. Inborn Errors of Metabolism. Trace Elements. Porphyrins. Purines and Nucleotides. Vitamins. Toxicology and Drug Monitoring. Neurogenic Amines and Secreting Tumours. Cancer. Subject Index.

1979 1,384 pages 0-444-80089-1 $79.00/Dfl. 170.00

Order these important books today!

ORDER FORM

Please send me:

____copy(ies) **Essentials of Clinical Biochemistry, Baron,** 0-444-00684-2, $26.95

(Distributed outside North America by Hodder & Stoughton Ltd., United Kingdom)

____copy(ies) **Chemical Diagnosis of Disease,** edited by **Brown/Mitchell/Young,** 0-444-80089-1, $79.00/Dfl. 170.00

Enclosed is my:

- ☐ personal check
- ☐ bank draft
- ☐ UNESCO coupons in the amount of ________

☐ Please bill me.

Please charge to:

- ☐ American Express
- ☐ VISA
- ☐ Master Card/Access/Eurocard (issuing bank No. __________.)

Account #__________ Expires________

Signature__________

Name__________

Address__________

Send to your bookseller, or
in the U.S.A. and Canada:
Elsevier Science Publishing Company, Inc.
52 Vanderbilt Avenue, New York, NY 10017

in all other countries:
Elsevier Science Publishers
P.O. Box 211, 1000 AE
Amsterdam, The Netherlands

ELSEVIER BIOMEDICAL

Book prices subject to change without notice. Printed in U.S.A. 7/82

Groundbreaking work in cancer treatment...

IMMUNOTHERAPY OF HUMAN CANCER

William D. Terry, M.D. and **Steven A. Rosenberg,** M.D., Ph.D., National Cancer Institute, National Institutes of Health, Bethesda, Maryland, editors

Dr. William D. Terry and Dr. Steven A. Rosenberg have compiled a unique text on the immunotherapy of human cancer which for the first time assembles in a single volume all modern clinical trials of immunotherapy. Oncologists, cancer researchers, immunologists and medical students will find this a valuable source of information on recent and current clinical trials, and on newer approaches to immunotherapy.

The volume contains 70 chapters dealing with immunotherapy for acute myelogenous leukemia, acute lymphocytic leukemia, non-Hodgkin's lymphoma, multiple myeloma, melanoma, and lung, breast, colorectal, genitourinary and gynecologic cancers. Results with interferon and other agents are also presented.

1982 520 pages 0-444-00614-1
$85.00 (Dfl.240.00 / US$ 111.50 outside North America)

EXCERPTA MEDICA

Order Form

Please send me ____ copy(ies) **Immunotherapy of Human Cancer,** Terry / Rosenberg, editors, 0-444-00614-1, $85.00 (Dfl.240.00 / $111.50 outside North America)

Enclosed is my:
☐ personal check ☐ bank draft
☐ UNESCO coupons in the amount of ____________
☐ Please bill me.

Please charge to:
☐ American Express ☐ VISA
☐ Master Card / Access / Eurocard
(issuing bank # ______________________)

Account # ______________________ Expires______

Signature ______________________

Name ______________________

Address ______________________
____
____

Send to your bookseller, or

in North America:
Elsevier Science Publishing Company, Inc.
52 Vanderbilt Avenue
New York, NY 10017

in the rest of the world:
Elsevier Science Publishers
P.O. Box 211
Amsterdam
The Netherlands

Price subject to change without notice.

Information for Authors

Manuscripts may be sent to the Editor-in-Chief: C.L. Meinert, The Johns Hopkins University, School of Hygiene and Public Health, Department of Epidemiology, 615 North Wolfe Street, Baltimore, Maryland 21205, USA; or to N. Tygstrup, Medical Department A, Rigshospitalet, 9 Blegdamsvej, DK-2100, Copenhagen, Denmark. Manuscripts are accepted with the understanding that they are original, unpublished works and are not being submitted elsewhere.

The Manuscript. Submit in quadruplicate, in English, double-spaced on one side of the paper only, and divided into the following sections: Title Page, Abstract (not more than 200 words), Key Words or Phrases (3 or 4), Introduction, Methods, Results, Discussion, References, Acknowledgments, Tables and Figures, and Figure Legends. The title page should include the names and affiliations of the authors, the complete address and telephone or telex number of the corresponding author, and an abbreviated title of not more than 45 characters and spaces. Footnotes in the text should be defined on the page on which they appear and numbered consecutively with superscript arabic numerals. Identify any mathematical variables that may be confused with numbers, e.g., lower case L and the numeral 1. In addition to full-length papers, short communications, providing a brief but complete account of a particular topic, may be submitted (Abstract not required).

Tables. Each table should appear on a separate page and be numbered with arabic numerals in order of references in the text. Use only horizontal rules. Table titles should be brief, with detailed information appearing as footnotes and identified in the table by superscript lowercase letters.

Illustrations. Submit one set of original figures along with photocopies to be included with each copy of the manuscript. Type legends double-spaced on a separate page. Letters, arrows, and symbols used in the figures should be identified in the legend and not on the figure itself. Label each original figure on the back with a paste-on label (do not write directly on it) indicating the number of the figure, the corresponding author's name, and the top of the figure. Do not mount figures or attach them with paper clips; enclosing them in an envelope or between paper will suffice. Hand-drawn artwork, and photocopies are not acceptable as originals. For line artwork, originals must be black-ink drawings of professional quality or glossy photographs. Simple line drawings will be reduced to fit within a space of approximately 5 × 3 in. (13 × 7½ cm); the size of the lettering must be such that, when reduced, it will be approximately 3/16 in. (2½ mm) in height. Photocopies included with each copy of the manuscript should have the necessary legends and footnotes in the position they are to appear when printed.

References. Identify references in the text by arabic numerals in parentheses, and number consecutively in the Reference Section in order of their mention. References should include: names of *all* contributing authors (last names first); title of article; title of journal (abbreviate according to the style of *Index Medicus*) or book; volume number; location and name of publishing company (books only); inclusive pages; year of publication. Examples:

Journal. Freund PA: Ethical problems in human experimentation. N Engl J Med 273:687–692, 1965

Book. Hill AB: Statistical Methods in Clinical Preventive Medicine. New York: Oxford University Press, 1962

Chapter in Book. Blackburn H: Multifactor preventive trials (MPT) in coronary heart disease. In: Trends in Epidemiology, Stewart GT, Ed. Springfield: Charles C. Thomas, 1972

Proofs and reprints. The corresponding author will receive proofs, which should be proofread and returned with the orignal manuscript within *72 hours* of receipt. Corrections are limited to printer's errors—no substantial author's changes will be made. Twenty-five reprints are supplied free of charge; additional reprints may be ordered at the price listed on the order form accompanying the proofs.

Page Charges. Page charges are assessed only for excessively long articles or for pages in supplemental issues of the Journal, and can be obtained upon request from the Editor or Publisher.

Book Reviews. Books for review are to be sent to the Book Review Editor, R. Sylvester, Institut Jules Bordet, Rue Héger—Bordet 1, 1000 Brussels, Belgium.

Copyright. Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher. This transfer will insure the widest possible dissemination of information under U.S. copyright law.

NIH LIBRARY
3 1496 01062 4495